

# Fejér-monotone hybrid steepest descent method for affinely constrained and composite convex minimization tasks<sup>‡</sup>

Konstantinos Slavakis<sup>1</sup>      Isao Yamada<sup>2</sup>

<sup>1</sup>University at Buffalo, The State University of New York, Dept. of Electrical Eng.,  
Buffalo, NY 14260-2500, USA

E-mail: kslavaki@buffalo.edu

<sup>2</sup>Tokyo Institute of Technology, Dept. of Communications & Computer Eng.,  
Tokyo 152-8550, Japan

E-mail: isao@sp.ce.titech.ac.jp

**Abstract.** This paper introduces the *Fejér-monotone hybrid steepest descent method* (FM-HSDM), a new member to the HSDM family of algorithms, for solving affinely constrained minimization tasks in real Hilbert spaces, where convex smooth and non-smooth losses compose the objective function. FM-HSDM offers sequences of estimates which converge weakly and, under certain hypotheses, strongly to solutions of the task at hand. Fixed-point theory, variational inequalities and affine-nonexpansive mappings are utilized to devise a scheme that accomodates affine constraints in a more versatile way than state-of-the-art primal-dual techniques and the alternating direction method of multipliers do. Recursions can be tuned to score low computational footprints, well-suited for large-scale optimization tasks, without compromising convergence guarantees. In contrast to its HSDM's precursors, FM-HSDM enjoys Fejér monotonicity, the step-size parameter stays constant across iterations to promote convergence speed-ups of the sequence of estimates to a minimizer, while only Lipschitzian continuity, and not strong monotonicity, of the derivative of the smooth-loss function is needed to ensure convergence. Results on the rate of convergence to an optimal point are also presented.

*Keywords:* Convex optimization, composite loss, Hilbert space, affine constraints, nonexpansive mapping, Fejér monotonicity, fixed-point theory, variational inequality.

## 1. Introduction

### 1.1. Problem and notation

**Problem 1.** This paper considers the following composite convex minimization task:

$$\min_{x \in \mathcal{A} \cap \mathcal{X}} f(x) + g(x), \quad (1)$$

<sup>‡</sup> Preliminary parts of this study can be found in [23, 24].

where  $\mathcal{X}$  is a real Hilbert space, the loss functions  $f, g$  belong to the class  $\Gamma_0(\mathcal{X})$  of all convex, proper, and lower-semicontinuous functions from  $\mathcal{X}$  to  $(-\infty, +\infty]$  [2, p. 132],  $f$  is everywhere (Fréchet) differentiable with  $L$ -Lipschitz-continuous derivative  $\nabla f$ , *i.e.*, there exists an  $L \in \mathbb{R}_{>0}$  such that (s.t.)  $\|\nabla f(x_1) - \nabla f(x_2)\| \leq L\|x_1 - x_2\|, \forall x_1, x_2 \in \mathcal{X}$ , and  $\mathcal{A}$  is a closed affine subset of  $\mathcal{X}$ . Throughout the manuscript, it is assumed that (1) indeed possesses a solution.  $\square$

Symbols  $\mathbb{Z}$  and  $\mathbb{R}$  stand for sets of all integer and real numbers, respectively. Moreover,  $\mathbb{Z}_{>0} := \{1, 2, \dots\} \subset \{0, 1, 2, \dots\} =: \mathbb{Z}_{\geq 0}$ , while  $\mathbb{R}_{>0} := (0, +\infty)$ . The algorithms of this paper are built on a real Hilbert space  $\mathcal{X}$ , equipped with an inner product  $\langle \cdot | \cdot \rangle$ , with elements denoted by lower-case letters, *e.g.*,  $x$ . In the special case where  $\mathcal{X}$  is finite dimensional, *i.e.*, Euclidean, elements of  $\mathcal{X}$  are denoted by boldfaced lower-case letters, *e.g.*,  $\mathbf{x}$ , while boldfaced upper-case letters are reserved for matrices, *e.g.*,  $\mathbf{Q}$ . Symbol  $\text{Id}$  denotes the identity mapping in  $\mathcal{X}$ , *i.e.*,  $\text{Id } x = x, \forall x \in \mathcal{X}$ . In the special case where  $\mathcal{X}$  is Euclidean,  $\text{Id}$  boils down to the identity matrix, denoted by  $\mathbf{I}$ . Vector/matrix transposition is denoted by the superscript  $\top$ . For  $g \in \Gamma_0(\mathcal{X})$ ,  $\partial g$  denotes the set-valued subdifferential operator which is defined as  $x \mapsto \partial g(x) := \{\xi \in \mathcal{X} \mid g(x) + \langle x' - x | \xi \rangle \leq g(x'), \forall x' \in \mathcal{X}\}$ .

Let  $\mathfrak{B}(\mathcal{X}, \mathcal{X}')$  denote all bounded linear operators from  $\mathcal{X}$  to  $\mathcal{X}'$  [17], and  $\mathfrak{B}(\mathcal{X}) := \mathfrak{B}(\mathcal{X}, \mathcal{X})$ . For  $Q \in \mathfrak{B}(\mathcal{X}, \mathcal{X}')$ ,  $\|Q\| < \infty$  stands for the norm of  $Q$ . Mapping  $Q^* \in \mathfrak{B}(\mathcal{X}', \mathcal{X})$  stands for the adjoint of  $Q \in \mathfrak{B}(\mathcal{X}, \mathcal{X}')$  [17]. In the case of matrices, the adjoint of a mapping  $\mathbf{Q}$  is nothing but the transpose  $\mathbf{Q}^\top$ . Mapping  $Q \in \mathfrak{B}(\mathcal{X})$  is called self-adjoint if  $Q^* = Q$ . In the case of a symmetric matrix  $\mathbf{Q}$ ,  $\lambda(\mathbf{Q})$  denotes an eigenvalue of  $\mathbf{Q}$ . Further,  $\|\mathbf{Q}\| = \sigma_{\max}(\mathbf{Q}) := \lambda_{\max}^{1/2}(\mathbf{Q}^\top \mathbf{Q})$  stands for the (spectral) norm of  $\mathbf{Q}$ , where  $\sigma_{\max}(\cdot) \in \mathbb{R}_{>0}$  denotes the maximum singular value and  $\lambda_{\max}(\cdot)$  the maximum eigenvalue of a matrix.

## 1.2. Background and contributions

The concise description in (1) can be unfolded in several ways to describe a large variety of convex composite minimization tasks, *e.g.*,

$$\min_{x \in \mathcal{A}} \ell(x) + \sum_{j=1}^J \mathbf{g}_j(H_j x - r_j), \quad (2)$$

where  $\{\mathcal{X}_j\}_{j=0}^J$  are real Hilbert spaces,  $\ell \in \Gamma_0(\mathcal{X}_0)$ ,  $\mathbf{g}_j \in \Gamma_0(\mathcal{X}_j)$ ,  $H_j \in \mathfrak{B}(\mathcal{X}_0, \mathcal{X}_j)$  and  $r_j \in \mathcal{X}_j$ ,  $j \in \{1, \dots, J\}$ . Moreover,  $\nabla \ell$  is  $L$ -Lipschitz continuous and  $\mathcal{A}$  is a closed affine subset of  $\mathcal{X}_0$ . Indeed, it can be verified that (2) can be recast as (1) via  $\mathcal{X} := \mathcal{X}_0 \times \mathcal{X}_1 \times \dots \times \mathcal{X}_J = \{x := (x^{(0)}, x^{(1)}, \dots, x^{(J)}) \mid x^{(j)} \in \mathcal{X}_j, \forall j \in \{0, 1, \dots, J\}\}$ ,  $f(x) := \ell(x^{(0)})$ ,  $g(x) := \sum_{j=1}^J \mathbf{g}_j(x^{(j)})$ , and the closed affine set  $\mathcal{A} := \{x \in \mathcal{X} \mid x^{(0)} \in \mathcal{A}, x^{(j)} = H_j x^{(0)} - r_j, \forall j \in \{1, \dots, J\}\}$ . Task (2), in the case where  $J = 2$ ,  $\mathcal{X} := \mathcal{X}_0 = \mathcal{X}_1$ ,  $H_1 = \text{Id}$ ,  $r_1 = r_2 = 0$ , and  $\mathcal{A} := \mathcal{X}$ , *i.e.*,  $\min_{x \in \mathcal{X}} [\ell(x) + \mathbf{g}_1(x) + \mathbf{g}_2(H_2 x)]$ , has been already studied via the primal-dual algorithmic framework [7, 9, 10]. Gradient  $\nabla \ell$ , proximal mappings (*cf.* Definition 6)  $\text{Prox}_{\mathbf{g}_1}$  and  $\text{Prox}_{\mathbf{g}_2^*} = \text{Id} - \text{Prox}_{\mathbf{g}_2}$  [2, Rem. 14.4,

p. 198], where  $\mathbf{g}_2^*$  stands for the (Fenchel) conjugate of  $\mathbf{g}_2$ , as well as adjoint  $H_2^*$  are utilized in a computationally efficient way to generate a sequence  $(x_n)_{n \in \mathbb{Z}_{\geq 0}} \subset \mathfrak{X}$ , which converges weakly (and under certain hypotheses, strongly) to a solution of the previous minimization task. Moreover, task (2), in the case where  $J = 2$ ,  $\mathfrak{X} := \mathfrak{X}_0 = \mathfrak{X}_1 = \mathfrak{X}_2$ ,  $H_1 = H_2 = \text{Id}$ ,  $r_1 = r_2 = 0$ , and  $\mathcal{A} := \mathcal{X}$ , *i.e.*,  $\min_{x \in \mathfrak{X}} [\ell(x) + \mathbf{g}_1(x) + \mathbf{g}_2(x)]$ , has also attracted attention in the context of the “three-term operator splitting” framework [6, 11]. As in [7, 9, 10],  $\nabla \ell$ ,  $\text{Prox}_{\mathbf{g}_1}$  and  $\text{Prox}_{\mathbf{g}_2}$  are employed via computationally efficient recursions to generate a sequence which converges weakly (and under certain hypotheses, strongly) to a solution of the minimization task at hand. All studies in [6, 7, 9–11] set  $\mathcal{A} := \mathfrak{X}$ . In the case of  $\mathcal{A} \subsetneq \mathfrak{X}$ , one can accommodate the affine constraint  $\mathcal{A}$  via the use of the indicator function  $\iota_{\mathcal{A}} [\iota_{\mathcal{A}}(x) := 0, \text{ if } x \in \mathcal{A}, \text{ and } \iota_{\mathcal{A}}(x) := +\infty, \text{ if } x \notin \mathcal{A}]$  and the additional loss  $\mathbf{g}_3 := \iota_{\mathcal{A}}$ . According to the previous discussion, such an accommodation entails the use of  $\text{Prox}_{\iota_{\mathcal{A}}} = P_{\mathcal{A}}$ , where  $P_{\mathcal{A}}$  denotes the metric projection mapping onto  $\mathcal{A}$ . Mapping  $P_{\mathcal{A}}$  may become computationally demanding, *e.g.*, in the case where  $\mathfrak{X}$  is a Euclidean space and the affine constraints are described by a matrix of large dimensions (*cf.* Fact 24), since computing  $P_{\mathcal{A}}$  necessitates the costly singular value decomposition of the matrix under query (*cf.* Example 25). Task (1) in the case where  $\mathcal{X}$  is a Euclidean space and  $\mathcal{A} := \{\mathbf{x} \in \mathcal{X} \mid \mathbf{a}^\top \mathbf{x} = 0\}$ , for some  $\mathbf{a} \in \mathcal{X} \setminus \{\mathbf{0}\}$ , was treated, within a stochastic setting, in [18].

The celebrated alternating direction method of multipliers (ADMM) [4, 5, 12, 13, 25] deals with the task

$$\min_{(x^{(1)}, x^{(2)}) \in \mathfrak{X}_1 \times \mathfrak{X}_2} \mathbf{g}_1(x^{(1)}) + \mathbf{g}_2(x^{(2)}) \quad (3a)$$

$$\text{s.to } H_1 x^{(1)} + H_2 x^{(2)} = r, \quad (3b)$$

where  $H_j \in \mathfrak{B}(\mathfrak{X}_j, \mathfrak{X}_0)$  and  $r \in \mathfrak{X}_0$ . Again, (3) can be recast as (1) under the following setting:  $\mathcal{X} := \mathfrak{X}_1 \times \mathfrak{X}_2 = \{x := (x^{(1)}, x^{(2)}) \mid x^{(1)} \in \mathfrak{X}_1, x^{(2)} \in \mathfrak{X}_2\}$ ,  $f(x) := 0$ ,  $g(x) := \mathbf{g}_1(x^{(1)}) + \mathbf{g}_2(x^{(2)})$ , and  $\mathcal{A} := \{x \in \mathcal{X} \mid H_1 x^{(1)} + H_2 x^{(2)} = r\}$ . Provided that mappings  $\lambda H_1^* H_1 + \partial \mathbf{g}_1$  and  $\lambda H_2^* H_2 + \partial \mathbf{g}_2$ , where  $\lambda \in \mathbb{R}_{>0}$  is a user-defined parameter, are maximal monotone, so that  $(\lambda H_1^* H_1 + \partial \mathbf{g}_1)^{-1}$  and  $(\lambda H_2^* H_2 + \partial \mathbf{g}_2)^{-1}$  exist, the recursive application of  $(\lambda H_1^* H_1 + \partial \mathbf{g}_1)^{-1}$  and  $(\lambda H_2^* H_2 + \partial \mathbf{g}_2)^{-1}$  generates a sequence which converges weakly to a solution of (3) [5, 25]. ADMM enjoys extremely wide popularity for minimization problems in Euclidean spaces [4], at the expense of the potentially costly computation of the inverse mappings  $(\lambda H_1^* H_1 + \partial \mathbf{g}_1)^{-1}$  and  $(\lambda H_2^* H_2 + \partial \mathbf{g}_2)^{-1}$ : there are cases where computing the previous inverse mappings entails solving a convex minimization subtask.

The motivation for the present paper is the algorithmic solution given in the distributed minimization context of [21, 22]: For a Euclidean  $\mathfrak{X}$ , and a collection of loss functions  $\{\ell_j, \mathbf{g}_j \in \Gamma_0(\mathfrak{X})\}_{j=1}^J$ , where  $\ell_j$  is everywhere differentiable with an  $L_j$ -Lipschitz continuous  $\nabla \ell_j$ ,  $\forall j \in \{1, \dots, J\}$ , nodes  $\mathcal{N}$  ( $|\mathcal{N}| = J$ ), connected by edges  $\mathcal{E}$  within a network/graph  $\mathcal{G} := (\mathcal{N}, \mathcal{E})$ , operate in parallel and cooperate to solve

$$\min_{(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(J)}) \in \mathfrak{X}^J} \sum_{j=1}^J \ell_j(\mathbf{x}^{(j)}) + \sum_{j=1}^J \mathbf{g}_j(\mathbf{x}^{(j)}) \quad (4a)$$

$$\text{s.to } \mathbf{x}^{(1)} = \dots = \mathbf{x}^{(J)}, \quad (4b)$$

Each node  $j \in \mathcal{N}$  operates only on the pair  $(\ell_j, \mathbf{g}_j)$  and communicates the information regarding its updates to its neighboring nodes to cooperatively solve (4), under the consensus constraint of (4b). Once again, (4) can be seen as a special case of (1) under the following considerations:  $\mathcal{X} := \mathfrak{X}^J$ ,  $f(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(J)}) := \sum_{j=1}^J \ell(\mathbf{x}^{(j)})$ ,  $g(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(J)}) := \sum_{j=1}^J \mathbf{g}(\mathbf{x}^{(j)})$ , and  $\mathcal{A} := \{(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(J)}) \in \mathcal{X} \mid \mathbf{x}^{(1)} = \dots = \mathbf{x}^{(J)}\}$ . Upon defining the  $J \times J$  mixing matrices  $\mathbf{W} = [w_{ij}]$ ,  $\tilde{\mathbf{W}} = [\tilde{w}_{ij}]$ , [22] introduced the following recursions to solve (4): For an arbitrarily fixed starting-point  $J \times \dim \mathfrak{X}$  matrix  $\mathbf{X}_0$ , as well as  $\mathbf{X}_{1/2} := \mathbf{W}\mathbf{X}_0 - \lambda \nabla f(\mathbf{X}_0)$  and  $\mathbf{X}_1 := \text{Prox}_{\lambda g}(\mathbf{X}_{1/2})$ , repeat for all  $n \in \mathbb{Z}_{\geq 0}$ , (i)  $\mathbf{X}_{n+3/2} := \mathbf{X}_{n+1/2} + \mathbf{W}\mathbf{X}_{n+1} - \tilde{\mathbf{W}}\mathbf{X}_n - \lambda[\nabla f(\mathbf{X}_{n+1}) - \nabla f(\mathbf{X}_n)]$ ; (ii)  $\mathbf{X}_{n+2} := \text{Prox}_{\lambda g}(\mathbf{X}_{n+3/2})$ . If (i)  $(i, j) \notin \mathcal{E} \Rightarrow w_{ij} = \tilde{w}_{ij} = 0$ , (ii)  $\mathbf{W}^\top = \mathbf{W}$ ,  $\tilde{\mathbf{W}}^\top = \tilde{\mathbf{W}}$ , (iii)  $\ker(\mathbf{W} - \tilde{\mathbf{W}}) = \text{span } \mathbf{1} \subset \ker(\mathbf{I} - \tilde{\mathbf{W}})$ , (iv)  $\tilde{\mathbf{W}} \succ \mathbf{0}$ , (v)  $(1/2)(\mathbf{I} + \tilde{\mathbf{W}}) \succeq \tilde{\mathbf{W}} \succeq \mathbf{W}$ , and (vi)  $\lambda \in (0, 2\lambda_{\min}(\tilde{\mathbf{W}})/\max_i L_i)$ , then the sequence  $(\mathbf{X}_n)_{n \in \mathbb{N}}$  converges to a matrix whose rows provide a solution to (4).

Aiming at the general problem of (1), this paper explores the directions established by the hybrid steepest descent method (HSDM) [26]. HSDM solves

$$\min_{x \in \text{Fix } T} f(x), \quad (5)$$

where  $\nabla f$  is strongly monotone [2], and  $\text{Fix } T \subset \mathcal{X}$  denotes the fixed-point set of a nonexpansive mapping  $T : \mathcal{X} \rightarrow \mathcal{X}$  (cf. Sec. 2). To solve (5), and for an arbitrarily fixed starting point  $x_0$ , HSDM generates the sequence

$$x_{n+1} := Tx_n - \lambda_n \nabla f(Tx_n), \quad (6)$$

which strongly converges to the *unique* minimizer of (5). To secure strong convergence, the sequence of step sizes  $(\lambda_n)_{n \in \mathbb{Z}_{\geq 0}} \subset \mathbb{R}_{\geq 0}$  is required to satisfy (i)  $\sum_{n \in \mathbb{Z}_{\geq 0}} \lambda_n = +\infty$ , (ii)  $\lim_{n \rightarrow \infty} \lambda_n = 0$ , and (iii)  $\sum_{n \in \mathbb{Z}_{\geq 0}} |\lambda_{n+1} - \lambda_n| < +\infty$ . Further, in the case where  $\mathcal{X}$  is Euclidean,  $\nabla f$  is not necessarily strongly monotone, and  $T$  is attracting nonexpansive with bounded  $\text{Fix } T$ , the requirements on the sequence  $(\lambda_n)_{n \in \mathbb{Z}_{\geq 0}}$  can be relaxed to (i)  $\sum_{n \in \mathbb{Z}_{\geq 0}} \lambda_n = +\infty$ , (ii)  $\sum_{n \in \mathbb{Z}_{\geq 0}} \lambda_n^2 < +\infty$  for achieving  $\lim_{n \rightarrow \infty} d_{\mathcal{X}}(\mathbf{x}_n, \text{Arg min}_{\text{Fix } T} f) = 0$ , where  $d_{\mathcal{X}}(\mathbf{x}_n, \text{Arg min}_{\text{Fix } T} f)$  stands for the (metric) distance of point  $\mathbf{x}_n$  from the set of minimizers of  $f$  within  $\text{Fix } T$  [19]. To speed up HSDM's convergence rate, conjugate-gradient-based variants were introduced in [14–16]. For example, for an arbitrarily fixed starting point  $x_0 \in \mathcal{X}$ , and  $d_0 := -\nabla f(x_0)$ , the following recursions (i)  $x_{n+1} := T(x_n + \mu \lambda_n d_n)$ ; (ii)  $d_{n+1} := -\nabla f(x_{n+1}) + \beta_{n+1} d_n$ , with  $\mu > 0$ ,  $\lambda_n \in (0, 1]$ ,  $\beta_n \in [0, \infty)$ , were introduced in [16]. If  $\mu \in (0, 2\eta/L^2)$ ,  $\lim_{n \rightarrow \infty} \beta_n = 0$ ,  $(\nabla f(x_n))_{n \in \mathbb{Z}_{\geq 0}}$  is bounded, and (i)  $\sum_{n \in \mathbb{Z}_{\geq 0}} \lambda_n = +\infty$ , (ii)  $\lim_{n \rightarrow \infty} \lambda_n = 0$ , (iii)  $\sum_{n \in \mathbb{Z}_{\geq 0}} |\lambda_{n+1} - \lambda_n| < +\infty$ , (iv)  $\lambda_n/\lambda_{n+1} \leq \sigma$ , ( $\sigma \geq 1$ ), then  $(x_n)_{n \in \mathbb{Z}_{\geq 0}}$  converges strongly to the unique minimizer of (5).

Driven by the similarity between the algorithmic solution of [21,22] and HSDM, this study introduces a new member to the HSDM family of algorithms, the *Fejér-monotone* (FM-)HSDM, for solving (1). Building around the remarkably simple recursion of (6)

and the powerful concept of a nonexpansive mapping, FM-HSDM's recursions offer sequences which converge weakly and, under certain hypotheses (uniform convexity of loss functions), strongly to a solution of (1); *cf.* Theorems 18 and 19. Fixed-point theory, variational inequalities and affine-nonexpansive mappings are utilized to accomodate the affine constraint  $\mathcal{A}$  in a more flexible way (see, *e.g.*, Proposition 10 and Example 25) than the usage of the indicator function and its associated metric-projection mapping that methods [6, 9–11] promote. Such flexibility is combined with the first-order information of  $f$  and the proximal mapping of  $g$  to build recursions of tunable complexity that can score low-computational-complexity footprints, well-suited for large-scale minimization tasks. FM-HSDM enjoys Fejér monotonicity, and in contrast to (6) as well as its conjugate-gradient-based variants [14–16], only Lipschitzian continuity, and not strong monotonicity, of the derivative of the smooth-part loss is needed to establish convergence of the sequence of estimates. Further, a constant step-size parameter is utilized to effect convergence speed-ups. Finally, as opposed to [14–16], the advocated scheme needs no boundedness assumptions on estimates or gradients to establish weak (or even strong) convergence of the sequence of estimates to a solution of (1). Results on the rate of convergence to an optimal point are also presented.

## 2. Affine-nonexpansive mappings and variational inequality problems

### 2.1. Fixed points and nonexpansive mappings

**Definition 2.** A self-adjoint mapping  $Q \in \mathfrak{B}(\mathcal{X})$  is called *positive* if  $\langle Qx | x \rangle \geq 0$ ,  $\forall x \in \mathcal{X}$  [17, § 9.3]. Moreover, the self-adjoint  $\Pi \in \mathfrak{B}(\mathcal{X})$  is called *strongly positive* if there exists  $\delta \in \mathbb{R}_{>0}$  s.t.  $\langle \Pi x | x \rangle \geq \delta \|x\|^2$ ,  $\forall x \in \mathcal{X}$ . In the context of matrices,  $\mathbf{Q}$  is positive iff  $\mathbf{Q}$  is positive semidefinite, *i.e.*,  $\mathbf{Q} \succeq \mathbf{0}$ . Moreover,  $\mathbf{\Pi}$  is strongly positive iff  $\mathbf{\Pi}$  is positive definite, *i.e.*,  $\mathbf{\Pi} \succ \mathbf{0}$ , and  $\delta$  in the previous definition can be taken to be  $\lambda_{\min}(\mathbf{\Pi})$ .  $\square$

For a strongly positive  $\Pi$ ,  $\langle \cdot | \cdot \rangle_{\Pi}$  stands for the inner product  $\langle x | x' \rangle_{\Pi} := \langle x | \Pi x' \rangle$ ,  $\forall (x, x') \in \mathcal{X}^2$ . For a function  $\varphi : \mathcal{X} \rightarrow \mathbb{R}$ ,  $\nabla \varphi$  and  $\nabla \varphi(x)$  stand for the (Gâteaux/Fréchet) derivative and gradient at  $x \in \mathcal{X}$ , respectively [2, § 2.6, p. 37]. Given  $Q \in \mathfrak{B}(\mathcal{X})$ ,  $\ker Q$  stands for the linear subspace  $\ker Q := \{x \in \mathcal{X} | Qx = 0\}$ . Moreover,  $\text{ran } Q$  denotes the linear subspace  $\text{ran } Q := Q\mathcal{X} := \{Qx | x \in \mathcal{X}\}$ . For the case of a matrix  $\mathbf{Q}$ ,  $\text{ran } \mathbf{Q}$  is the linear subspace spanned by the columns of  $\mathbf{Q}$ . Finally, the orthogonal complement of a linear subspace is denoted by the superscript  $\perp$ .

**Definition 3.** The *fixed-point set* of a mapping  $T : \mathcal{X} \rightarrow \mathcal{X}$  is defined as the set  $\text{Fix } T := \{x \in \mathcal{X} | Tx = x\}$ .  $\square$

**Definition 4.** Mapping  $T : \mathcal{X} \rightarrow \mathcal{X}$  is called

- (i) *Nonexpansive*, if  $\|Tx - Tx'\| \leq \|x - x'\|$ ,  $\forall (x, x') \in \mathcal{X}^2$ .
- (ii) *Firmly nonexpansive*, if  $\|Tx - Tx'\|^2 + \|(\text{Id} - T)x - (\text{Id} - T)x'\|^2 \leq \|x - x'\|^2$ ,  $\forall (x, x') \in \mathcal{X}^2$ .

- (iii)  $\alpha$ -averaged (nonexpansive), if there exist an  $\alpha \in (0, 1)$  and a nonexpansive mapping  $R : \mathcal{X} \rightarrow \mathcal{X}$  s.t.  $T = \alpha R + (1 - \alpha) \text{Id}$ . It can be easily verified that  $T$  is nonexpansive with  $\text{Fix } R = \text{Fix } T$ .  $\square$

**Fact 5** ([2, Cor. 4.15, p. 63]). The fixed-point set  $\text{Fix } T$  of a nonexpansive mapping  $T$  is closed and convex.  $\square$

**Definition 6.** Given  $f \in \Gamma_0(\mathcal{X})$  and  $\gamma \in \mathbb{R}_{>0}$ , the *proximal* mapping  $\text{Prox}_{\gamma f}$  is defined as  $\text{Prox}_{\gamma f} : \mathcal{X} \rightarrow \mathcal{X} : x \mapsto \arg \min_{z \in \mathcal{X}} (\gamma f(z) + \frac{1}{2} \|x - z\|^2)$ .  $\square$

**Example 7.**

- (i) [2, Prop. 4.8, p. 61] Given a non-empty closed convex set  $\mathcal{C} \subset \mathcal{X}$ , the *metric projection mapping onto  $\mathcal{C}$* , defined as  $P_{\mathcal{C}} : \mathcal{X} \rightarrow \mathcal{C} : x \mapsto P_{\mathcal{C}}x$ , with  $P_{\mathcal{C}}x$  being the unique minimizer of  $\min_{z \in \mathcal{C}} \|x - z\|$ , is firmly nonexpansive with  $\text{Fix } P_{\mathcal{C}} = \mathcal{C}$ .
- (ii) [2, Prop. 12.27, p. 176] Given  $f \in \Gamma_0(\mathcal{X})$  and  $\gamma \in \mathbb{R}_{>0}$ , the proximal mapping  $\text{Prox}_{\gamma f}$  is firmly nonexpansive with  $\text{Fix } \text{Prox}_{\gamma f} = \arg \min f$ .
- (iii) [2, Prop. 4.2, p. 60]  $T$  is firmly nonexpansive iff  $\text{Id} - T$  is firmly nonexpansive iff  $T$  is  $(1/2)$ -averaged iff  $2T - \text{Id}$  is nonexpansive.
- (iv) [8, Prop. 2.2], [19, Thm. 3(b)]. Let  $\{T_j\}_{j=1}^J$  be a finite family ( $J \in \mathbb{Z}_{>0}$ ) of nonexpansive mappings from  $\mathcal{X}$  to  $\mathcal{X}$ , and  $\{\omega_j\}_{j=1}^J$  be real numbers in  $(0, 1]$  s.t.  $\sum_{j=1}^J \omega_j = 1$ . Then,  $T := \sum_{j=1}^J \omega_j T_j$  is nonexpansive. If  $\cap_{j=1}^J \text{Fix } T_j \neq \emptyset$ , then  $\text{Fix } T = \cap_{j=1}^J \text{Fix } T_j$ . Further, consider real numbers  $\{\alpha_j\}_{j=1}^J \subset (0, 1)$  s.t.  $T_j$  is  $\alpha_j$ -averaged,  $\forall j$ . Define  $\alpha := \sum_{j=1}^J \omega_j \alpha_j$ . Then,  $T$  is  $\alpha$ -averaged.
- (v) [8, Prop. 2.5], [19, Thm. 3(b)] Let  $\{T_j\}_{j=1}^J$  be a finite family ( $J \in \mathbb{Z}_{>0}$ ) of nonexpansive mappings from  $\mathcal{X}$  to  $\mathcal{X}$ . Then, mapping  $T := T_1 T_2 \cdots T_J$  is nonexpansive. If  $\cap_{j=1}^J \text{Fix } T_j \neq \emptyset$ , then  $\text{Fix } T = \cap_{j=1}^J \text{Fix } T_j$ . Further, consider real numbers  $\{\alpha_j\}_{j=1}^J \subset (0, 1)$  s.t.  $T_j$  is  $\alpha_j$ -averaged,  $\forall j$ . Define

$$\alpha := \frac{1}{1 + \frac{1}{\sum_{j=1}^J \frac{1}{1 - \alpha_j}}}.$$

Then,  $T$  is  $\alpha$ -averaged.  $\square$

In what follows, function  $f \in \Gamma_0(\mathcal{X})$  is considered to have an  $L$ -Lipschitz continuous  $\nabla f$  with  $\text{dom } \nabla f = \mathcal{X}$ . By [2, Prop. 16.3(i), p. 224], the previous condition leads to  $\text{dom } f = \mathcal{X}$ , which further implies by [2, Cor. 16.38(iii), p. 234] that  $\partial(f + g) = \nabla f + \partial g$ .

## 2.2. Affine-nonexpansive mappings

**Definition 8** ([2, p. 3]). A mapping  $T : \mathcal{X} \rightarrow \mathcal{X}$  is called *affine* if there exist a linear mapping  $Q : \mathcal{X} \rightarrow \mathcal{X}$  and a  $\pi \in \mathcal{X}$  s.t.  $Tx = Qx + \pi$ ,  $\forall x \in \mathcal{X}$ .  $\square$

**Fact 9** ([2, Ex. 4.4, p. 72]). Consider the affine mapping  $Tx = Qx + \pi$ ,  $\forall x \in \mathcal{X}$ , with  $Q$  being linear and  $\pi \in \mathcal{X}$ . Then,  $T$  is nonexpansive iff  $\|Q\| \leq 1$ .  $\square$

Define now the following special class of affine-nonexpansive mappings:

$$\mathfrak{T} := \left\{ T : \mathcal{X} \rightarrow \mathcal{X} \left| \begin{array}{l} Tx = Qx + \pi, \forall x \in \mathcal{X}; \\ Q \in \mathfrak{B}(\mathcal{X}); \pi \in \mathcal{X}; \\ \|Q\| \leq 1, Q \text{ is positive} \end{array} \right. \right\}. \quad (7)$$

Interestingly, as the following proposition shows,  $\mathfrak{T}$  is closed with respect to convex combination and a specific form of operator composition.

**Proposition 10.** Consider a finite family  $\{T_j\}_{j=1}^J$  ( $J \in \mathbb{Z}_{>0}$ ) of members of  $\mathfrak{T}$ .

- (i) For any set of weights  $\{\omega_j\}_{j=1}^J$  s.t.  $\omega_j \in (0, 1]$  and  $\sum_{j=1}^J \omega_j = 1$ , mapping  $\sum_{j=1}^J \omega_j T_j x \in \mathfrak{T}$ .
- (ii) Given also  $T_0 \in \mathfrak{T}$ , the composition

$$\begin{aligned} T_J T_{J-1} \cdots T_1 T_0 T_1 \cdots T_{J-1} T_J x &= Q_J Q_{J-1} \cdots Q_1 Q_0 Q_1 \cdots Q_{J-1} Q_J x \\ &\quad + \sum_{j=1}^J Q_J Q_{J-1} \cdots Q_1 Q_0 Q_1 \cdots Q_{j-1} \pi_j \\ &\quad + \sum_{j=1}^J Q_J Q_{J-1} \cdots Q_j \pi_{j-1} + \pi_J, \quad \forall x \in \mathcal{X}, \end{aligned}$$

satisfies  $T_J T_{J-1} \cdots T_1 T_0 T_1 \cdots T_{J-1} T_J \in \mathfrak{T}$ . □

*Proof.* The proof of Proposition 10(i) follows easily from the definition of  $\sum_{j=1}^J \omega_j T_j x$  and  $\|\sum_j \omega_j Q_j\| \leq \sum_j \omega_j \|Q_j\| \leq \sum_j \omega_j = 1$ . The formula appearing in Proposition 10(ii) can be deduced by mathematical induction on  $J$ . Further,  $Q_J Q_{J-1} \cdots Q_1 Q_0 Q_1 \cdots Q_{J-1} Q_J$  is self-adjoint, and its positiveness follows from the fundamental observation that  $\forall x \in \mathcal{X}$ ,  $\langle Q_J Q_{J-1} \cdots Q_1 Q_0 Q_1 \cdots Q_{J-1} Q_J x \mid x \rangle = \langle Q_0(Q_1 \cdots Q_{J-1} Q_J x) \mid Q_1 \cdots Q_{J-1} Q_J x \rangle \geq 0$ , due to the positiveness of  $Q_0$ . Finally, the claim of Proposition 10(ii) is established by  $\|Q_J \cdots Q_1 Q_0 Q_1 \cdots Q_J\| \leq \|Q_0\| \prod_{j=1}^J \|Q_j\|^2 \leq 1$ . □

**Proposition 11.** Given the closed affine set  $\mathcal{A} \subset \mathcal{X}$ , define the following family of mappings:

$$\mathfrak{T}_{\mathcal{A}} := \{T \in \mathfrak{T} \mid \text{Fix } T = \mathcal{A}\}. \quad (8)$$

Then,  $\mathfrak{T}_{\mathcal{A}}$  is non-empty. More specifically, the metric projection mapping  $P_{\mathcal{A}}$  onto  $\mathcal{A}$  belongs to  $\mathfrak{T}_{\mathcal{A}}$ . □

*Proof.* For an arbitrarily fixed  $w \in \mathcal{A}$ , the set  $\mathcal{V} := \mathcal{A} - w$  is a closed linear subspace ( $\mathcal{V}$  does not depend on the choice of  $w$ ). By [2, Prop. 3.17, p. 47], the metric projection mapping onto  $\mathcal{A}$  becomes  $P_{\mathcal{A}} = P_{\mathcal{V}} + (w - P_{\mathcal{V}}w)$ , where  $P_{\mathcal{V}}$  is the metric projection mapping onto  $\mathcal{V}$ . Now, by [2, Cor. 3.22, p. 49], it can be verified that  $P_{\mathcal{V}}$  is linear and positive, with  $\|P_{\mathcal{V}}\| = 1$ . Moreover, due to Example 7(i),  $\text{Fix } P_{\mathcal{A}} = \mathcal{A}$ . Hence,  $P_{\mathcal{A}} \in \mathfrak{T}_{\mathcal{A}}$ . □

It can be verified that the fixed-point set  $\text{Fix } T$  of an affine mapping  $T$  is affine. However, more can be said about the members of  $\mathfrak{T}_{\mathcal{A}}$ .

**Proposition 12.** For any  $T \in \mathfrak{T}_{\mathcal{A}}$ ,

$$\mathcal{A} = \text{Fix } T = \ker(\text{Id} - Q) + w_* = \ker U + w_*,$$

where  $w_*$  is any element of  $\mathcal{A}$ , and  $U$  is the *positive square root* of  $\text{Id} - Q$ , i.e., the (unique) positive operator which satisfies  $U^2 = \text{Id} - Q$  [17, Thm. 9.4-2, p. 476].  $\square$

*Proof.* Since  $\|Q\| = \sup_{x \in \mathcal{X} \setminus \{0\}} |\langle Qx \mid x \rangle| / \|x\|^2$  [17, Thm. 9.2-2, p. 466] and  $\|Q\| \leq 1$ , it can be easily verified that  $\forall x \in \mathcal{X}$ ,  $\langle (\text{Id} - Q)x \mid x \rangle = \|x\|^2 - \langle Qx \mid x \rangle \geq \|x\|^2 - \|Q\| \cdot \|x\|^2 \geq \|x\|^2 - \|x\|^2 = 0$ , i.e.,  $\text{Id} - Q$  is positive. Interestingly, the positiveness of  $Q$  suggests that  $\forall x \in \mathcal{X}$ ,  $\langle (\text{Id} - Q)x \mid x \rangle = \|x\|^2 - \langle Qx \mid x \rangle \leq \|x\|^2$ , which implies, via [17, Thm. 9.2-2, p. 466], that  $\|\text{Id} - Q\| \leq 1$ . Moreover, by the definition of  $T$ , it follows that for any arbitrarily fixed  $w_* \in \text{Fix } T$ ,

$$\begin{aligned} \text{Fix } T &= \{x \mid Tx = x\} = \{x \mid (\text{Id} - T)x = 0\} \\ &= \{x \mid (\text{Id} - Q)x = \pi\} = \{x \mid (\text{Id} - Q)x = (\text{Id} - Q)w_*\} \\ &= \{x \mid (\text{Id} - Q)(x - w_*) = 0\} = \{x' + w_* \mid (\text{Id} - Q)x' = 0\} \\ &= \ker(\text{Id} - Q) + w_*. \end{aligned}$$

Finally, the characterization  $\text{Fix } T = \ker U + w_*$  follows from the previous arguments and  $x' \in \ker(\text{Id} - Q) \Leftrightarrow (\text{Id} - Q)x' = 0 \Rightarrow U^2x' = 0 \Rightarrow U^*Ux' = 0 \Rightarrow \langle x' \mid U^*Ux' \rangle = \langle Ux' \mid Ux' \rangle = \|Ux'\|^2 = 0 \Rightarrow Ux' = 0 \Leftrightarrow x' \in \ker U \Rightarrow U^2x' = 0 \Rightarrow (\text{Id} - Q)x' = 0 \Leftrightarrow x' \in \ker(\text{Id} - Q)$ , which establishes  $\ker(\text{Id} - Q) = \ker U$ .  $\square$

Several examples of  $\mathfrak{T}_{\mathcal{A}}$  members playing important roles in convex minimization tasks can be found in Appendix A.

### 2.3. Variational inequality problems

**Definition 13** (Variational inequality problem). For a nonexpansive mapping  $T : \mathcal{X} \rightarrow \mathcal{X}$ , point  $x_* \in \text{Fix } T$  is said to solve the variational inequality problem  $\text{VIP}(\nabla f + \partial g, \text{Fix } T)$  if there exists  $\xi_* \in \partial g(x_*)$  s.t.  $\forall y \in \text{Fix } T$ ,  $\langle y - x_* \mid \nabla f(x_*) + \xi_* \rangle \geq 0$ .  $\square$

**Fact 14** ([2, Prop. 26.5(vi), p. 383]). Consider a mapping  $T \in \mathfrak{T}_{\mathcal{A}}$  (recall  $\text{Fix } T = \mathcal{A}$ ), and assume that one of the following holds:

- (i)  $0 \in \text{sri}(\mathcal{A} - \text{dom}(f + g))$  (cf. [2, Prop. 6.19, p. 95] for special cases);
- (ii)  $\mathcal{X}$  is Euclidean and  $\mathcal{A} \cap \text{ri}[\text{dom}(f + g)] \neq \emptyset$ .

Then, point  $x_*$  solves  $\text{VIP}(\nabla f + \partial g, \text{Fix } T)$  iff  $x_* \in \arg \min_{x \in \text{Fix } T} [f(x) + g(x)]$ .  $\square$

**Proposition 15.** Given the closed affine set  $\mathcal{A} \subset \mathcal{X}$ , consider any  $T \in \mathfrak{T}_{\mathcal{A}}$  (cf. Proposition 12). If  $U$  stands for the square root of the linear operator  $\text{Id} - Q$  in the



description of  $T$  (cf. Definition 8), let  $\overline{\text{ran}} U$  denote the closure (in the strong topology) of the range of  $U$ . Then,

$$\begin{aligned} x_* \text{ solves } \text{VIP}(\nabla f + \partial g, \text{Fix } T) \\ \Leftrightarrow x_* \in \Gamma_* := \{x \in \text{Fix } T \mid [\nabla f(x) + \partial g(x)] \cap \overline{\text{ran}} U \neq \emptyset\}. \end{aligned} \quad (9a)$$

Moreover, for an arbitrarily fixed  $\lambda \in \mathbb{R} \setminus \{0\}$ , define the subset

$$\mathcal{O}_\lambda := \{(x, v) \in \text{Fix } T \times \mathcal{X} \mid -\frac{1}{\lambda} Uv \in \nabla f(x) + \partial g(x)\}. \quad (9b)$$

Then,

$$(x_*, v_*) \in \mathcal{O}_\lambda \Rightarrow x_* \text{ solves } \text{VIP}(\nabla f + \partial g, \text{Fix } T). \quad (9c)$$

Further, in the case where  $\mathcal{X}$  is Euclidean,

$$\mathbf{x}_* \text{ solves } \text{VIP}(\nabla f + \partial g, \text{Fix } T) \Leftrightarrow \exists \mathbf{v}_* \in \mathcal{X} \text{ s.t. } (\mathbf{x}_*, \mathbf{v}_*) \in \mathcal{O}_\lambda. \quad (9d)$$

□

*Proof.* First, recall that  $(\ker U)^\perp = \overline{\text{ran}} U^* = \overline{\text{ran}} U$  [2, Fact 2.18(iii), p. 32]. According to Definition 13,

$$\begin{aligned} x_* \text{ solves } \text{VIP}(\nabla f + \partial g, \text{Fix } T) \\ \Leftrightarrow x_* \in \text{Fix } T \text{ and } \exists \xi_* \in \partial g(x_*) \text{ s.t. } \forall y \in \text{Fix } T, \langle y - x_* \mid \nabla f(x_*) + \xi_* \rangle \geq 0 \\ \Leftrightarrow x_* \in \text{Fix } T \text{ and } \exists \xi_* \in \partial g(x_*) \text{ s.t. } \forall z \in \ker U, \langle z \mid \nabla f(x_*) + \xi_* \rangle \geq 0 \quad (10a) \\ \Leftrightarrow x_* \in \text{Fix } T \text{ and } \exists \xi_* \in \partial g(x_*) \text{ s.t. } \forall z \in \ker U, \langle z \mid \nabla f(x_*) + \xi_* \rangle \leq 0 \quad (10b) \\ \Leftrightarrow x_* \in \text{Fix } T \text{ and } \exists \xi_* \in \partial g(x_*) \text{ s.t. } \forall z \in \ker U, \langle z \mid \nabla f(x_*) + \xi_* \rangle = 0 \\ \Leftrightarrow x_* \in \text{Fix } T \text{ and } \exists \xi_* \in \partial g(x_*) \text{ s.t. } \nabla f(x_*) + \xi_* \in (\ker U)^\perp = \overline{\text{ran}} U \\ \Leftrightarrow x_* \in \text{Fix } T \text{ and } [\nabla f(x_*) + \partial g(x_*)] \cap \overline{\text{ran}} U \neq \emptyset \\ \Leftrightarrow x_* \in \Gamma_*, \end{aligned} \quad (10c)$$

which establishes (9a). Notice that Proposition 12 was used in (10a) and  $z \in \ker U \Leftrightarrow -z \in \ker U$  in (10b).

Moreover,

$$\begin{aligned} (x_*, v_*) \in \mathcal{O}_\lambda \\ \Leftrightarrow x_* \in \text{Fix } T \text{ and } U \left(-\frac{v_*}{\lambda}\right) \in \nabla f(x_*) + \partial g(x_*) \\ \Leftrightarrow x_* \in \text{Fix } T \text{ and } \exists v'_* \in \mathcal{X} \text{ s.t. } Uv'_* \in [\nabla f(x_*) + \partial g(x_*)] \cap \text{ran } U \quad (v'_* = -\frac{v_*}{\lambda}) \\ \Leftrightarrow x_* \in \text{Fix } T \text{ and } [\nabla f(x_*) + \partial g(x_*)] \cap \text{ran } U \neq \emptyset \\ \Rightarrow x_* \in \text{Fix } T \text{ and } [\nabla f(x_*) + \partial g(x_*)] \cap \overline{\text{ran}} U \neq \emptyset \quad (11a) \\ \Leftrightarrow x_* \in \Gamma_*, \end{aligned}$$

which establishes (9c) via (9a).

In the case where  $\mathcal{X}$  is Euclidean, (9d) is established by the well-known fact  $\overline{\text{ran}} U = \text{ran } U$  [17, Thm. 2.4-3, p. 74], which turns “ $\Rightarrow$ ” into “ $\Leftrightarrow$ ” in (11a). □

### 3. Algorithm and convergence analysis

**Definition 16** ([2, (10.2), p. 144]). A proper convex function  $h : \mathcal{X} \rightarrow (-\infty, +\infty]$  is called *uniformly convex* on a non-empty subset  $\mathcal{S}$  of  $\text{dom } h$ , if there exists an increasing function  $\varphi_{\mathcal{S}} : [0, +\infty] \rightarrow [0, +\infty]$ , which vanishes only at 0, s.t.  $\forall x, x' \in \mathcal{S}$  and  $\forall \mu \in (0, 1)$ ,

$$h(\mu x + (1 - \mu)x') + \mu(1 - \mu)\varphi_{\mathcal{S}}(\|x - x'\|) \leq \mu h(x) + (1 - \mu)h(x').$$

In the case where  $\mathcal{S} := \text{dom } h$  and  $\varphi_{\mathcal{S}} := (\beta_{\mathcal{S}}/2)(\cdot)^2$ , for some  $\beta_{\mathcal{S}} \in \mathbb{R}_{>0}$ , then  $h$  is called *strongly convex* with constant  $\beta_{\mathcal{S}}$ . Moreover, “strong convexity”  $\Rightarrow$  “uniform convexity”  $\Rightarrow$  “strict convexity.”  $\square$

**Assumption 17.**

- (i) Function  $f$  is uniformly convex on every non-empty bounded subset of  $\mathcal{X}$ .
- (ii) Function  $g$  is uniformly convex on every non-empty bounded subset of  $\text{dom } \partial g$ .

$\square$

For any  $T \in \mathfrak{T}_{\mathcal{A}}$  and any  $\alpha \in (0, 1)$ , define the  $\alpha$ -averaged mapping

$$T_{\alpha}x := [\alpha T + (1 - \alpha)\text{Id}]x = Q_{\alpha}x + \alpha\pi, \quad (12)$$

where  $Q_{\alpha} := \alpha Q + (1 - \alpha)\text{Id}$ .

**Theorem 18.** Consider  $f, g \in \Gamma_0(\mathcal{X})$ , with  $L$  being the Lipschitz-continuity constant of  $\nabla f$ . Moreover, given the closed affine set  $\mathcal{A}$ , consider any  $T \in \mathfrak{T}_{\mathcal{A}}$ . For  $\lambda \in \mathbb{R}_{>0}$ , an arbitrarily fixed  $x_0 \in \mathcal{X}$ , and for all  $n \in \mathbb{Z}_{\geq 0}$ , the *Fejér-monotone hybrid steepest descent method* (FM-HSDM) is stated as follows:

$$x_{1/2} := T_{\alpha}x_0 - \lambda \nabla f(x_0), \quad (13a)$$

$$x_1 := \text{Prox}_{\lambda g}(x_{1/2}), \quad (13b)$$

$$x_{n+3/2} := x_{n+1/2} - [T_{\alpha}x_n - \lambda \nabla f(x_n)] + [Tx_{n+1} - \lambda \nabla f(x_{n+1})], \quad (13c)$$

$$x_{n+2} := \text{Prox}_{\lambda g}(x_{n+3/2}). \quad (13d)$$

In the case where the non-smooth part of the composite loss becomes zero, *i.e.*,  $g := 0$ , then (13) takes the special form

$$x_{1/2} := T_{\alpha}x_0 - \lambda \nabla f(x_0), \quad (14a)$$

$$x_1 := x_{1/2}, \quad (14b)$$

$$x_{n+3/2} := x_{n+1/2} - [T_{\alpha}x_n - \lambda \nabla f(x_n)] + [Tx_{n+1} - \lambda \nabla f(x_{n+1})], \quad (14c)$$

$$x_{n+2} := x_{n+3/2}. \quad (14d)$$

Further, in the case where  $f = 0$ , recursions take the form

$$x_{1/2} := T_{\alpha}x_0, \quad (15a)$$

$$x_1 := \text{Prox}_{\lambda g}(x_{1/2}), \quad (15b)$$

$$x_{n+3/2} := x_{n+1/2} - T_\alpha x_n + T x_{n+1}, \quad (15c)$$

$$x_{n+2} := \text{Prox}_{\lambda g}(x_{n+3/2}). \quad (15d)$$

Consider  $\alpha \in [0.5, 1)$  and  $\lambda \in (0, 2(1 - \alpha)/L)$ . Then, the following hold true.

- (i) For any sequence  $(x_n)_{n \in \mathbb{Z}_{\geq 0}}$  chosen from (13), (14) and (15), there exist a sequence  $(v_n)_{n \in \mathbb{Z}_{\geq 0}} \subset \mathcal{X}$  and a strongly positive operator  $\Theta : \mathcal{X}^2 \rightarrow \mathcal{X}^2$  s.t. sequence  $(y_n := (x_n, v_n))_{n \in \mathbb{Z}_{>0} \setminus \{1\}}$  is Fejér monotone [2, Def. 5.1, p. 75] w.r.t.  $\mathcal{O}_\lambda$  of Proposition 15 in the Hilbert space  $(\mathcal{X}^2, \langle \cdot | \cdot \rangle_\Theta)$ , i.e.,

$$\|(x_{n+1}, v_{n+1}) - (x_*, v_*)\|_\Theta \leq \|(x_n, v_n) - (x_*, v_*)\|_\Theta, \quad \forall (x_*, v_*) \in \mathcal{O}_\lambda.$$

- (ii) Sequences  $(x_n)_{n \in \mathbb{Z}_{\geq 0}}$  of (13), (14) and (15) converge weakly to points that solve  $\text{VIP}(\nabla f + \partial g, \text{Fix } T)$ ,  $\text{VIP}(\nabla f, \text{Fix } T)$  and  $\text{VIP}(\partial g, \text{Fix } T)$ , respectively.
- (iii) Additionally, if either Assumption 17(i) or Assumption 17(ii) hold true, then sequences  $(x_n)_{n \in \mathbb{Z}_{\geq 0}}$  of (13), (14) and (15) converge strongly to points that solve  $\text{VIP}(\nabla f + \partial g, \text{Fix } T)$ ,  $\text{VIP}(\nabla f, \text{Fix } T)$  and  $\text{VIP}(\partial g, \text{Fix } T)$ , respectively.  $\square$

*Proof.* Only the proof regarding sequence  $(x_n)_{n \in \mathbb{Z}_{\geq 0}}$  of (13) is provided here. The proofs regarding sequences (14) and (15) become special cases of the following one under  $g = 0$  and  $f = 0$ , respectively.

(i) By (13c),

$$x_{n+3/2} - x_{n+1/2} = T x_{n+1} - T_\alpha x_n - \lambda [\nabla f(x_{n+1}) - \nabla f(x_n)]. \quad (16)$$

Since  $z = \text{Prox}_{\lambda g}(y) \Leftrightarrow (\exists \xi \in \partial g(z) \text{ s.t. } z + \lambda \xi = y)$ , then

$$\exists \xi_{n+2} \in \partial g(x_{n+2}) \quad (17)$$

s.t.  $x_{n+3/2} = x_{n+2} + \lambda \xi_{n+2}$  and thus  $\exists \xi_{n+1} \in \partial g(x_{n+1})$  s.t.  $x_{n+1/2} = x_{n+1} + \lambda \xi_{n+1}$ . Incorporating the previous equations in (16) yields that  $\forall n \in \mathbb{Z}_{\geq 0}$ ,

$$\begin{aligned} x_1 &= T_\alpha x_0 - \lambda [\nabla f(x_0) + \xi_1], \\ x_{n+2} - x_{n+1} &= T x_{n+1} - T_\alpha x_n - \lambda [\nabla f(x_{n+1}) + \xi_{n+2}] + \lambda [\nabla f(x_n) + \xi_{n+1}]. \end{aligned} \quad (18)$$

Moreover, adding consecutive equations of (18) results into the following fact:  $\forall n \geq 2$ ,

$$\begin{aligned} x_{n+1} &= T x_n - \sum_{\nu=1}^{n-1} (T_\alpha - T) x_\nu - \lambda [\nabla f(x_n) + \xi_{n+1}] \\ &= T x_n - \sum_{\nu=1}^{n+1} (T_\alpha - T) x_\nu + (T_\alpha - T) x_n + (T_\alpha - T) x_{n+1} - \lambda [\nabla f(x_n) + \xi_{n+1}] \\ &= 2T_\alpha x_{n+1} - T x_{n+1} + (T_\alpha x_n - T_\alpha x_{n+1}) - \sum_{\nu=1}^{n+1} (T_\alpha - T) x_\nu - \lambda [\nabla f(x_n) + \xi_{n+1}]. \end{aligned}$$

Consequently,

$$(\text{Id} + T - 2T_\alpha) x_{n+1} + (T_\alpha x_{n+1} - T_\alpha x_n)$$

$$\begin{aligned}
&= (1 - 2\alpha)(T - \text{Id})x_{n+1} + Q_\alpha(x_{n+1} - x_n) \\
&= -\sum_{\nu=1}^{n+1} (T_\alpha - T)x_\nu - \lambda[\nabla f(x_n) + \xi_{n+1}] , \tag{19}
\end{aligned}$$

where the first equation is due to (12).

Choose arbitrarily a  $w_* \in \text{Fix } T$ , *i.e.*,  $(\text{Id} - T)w_* = 0$ . Then,

$$\begin{aligned}
(T_\alpha - T)x_\nu &= (1 - \alpha)(\text{Id} - T)x_\nu \\
&= (1 - \alpha)[(\text{Id} - T)x_\nu - (\text{Id} - T)w_*] \\
&= (1 - \alpha)(\text{Id} - Q)(x_\nu - w_*) .
\end{aligned}$$

Define also

$$v_{n+1} := (1 - \alpha) \sum_{\nu=1}^{n+1} U(x_\nu - w_*) .$$

Point  $v_{n+1}$  does not depend on the choice of the fixed point  $w_*$ . Indeed, by Proposition 12, it can be verified that for any  $w_\# \in \text{Fix } T$ ,  $w_\# - w_* \in \ker U$ , and that

$$\begin{aligned}
v_{n+1} &= (1 - \alpha) \sum_{\nu=1}^{n+1} U(x_\nu - w_\# + w_\# - w_*) \\
&= (1 - \alpha) \sum_{\nu=1}^{n+1} [U(x_\nu - w_\#) + U(w_\# - w_*)] \\
&= (1 - \alpha) \sum_{\nu=1}^{n+1} U(x_\nu - w_\#) . \tag{20}
\end{aligned}$$

Moreover,

$$\begin{aligned}
v_{n+1} - v_n &= (1 - \alpha) \sum_{\nu=1}^{n+1} U(x_\nu - w_*) - (1 - \alpha) \sum_{\nu=1}^n U(x_\nu - w_*) \\
&= (1 - \alpha)U(x_{n+1} - w_*) , \quad \forall w_* \in \text{Fix } T , \tag{21}
\end{aligned}$$

and

$$\begin{aligned}
-\sum_{\nu=1}^{n+1} (T_\alpha - T)x_\nu &= -(1 - \alpha) \sum_{\nu=1}^{n+1} (\text{Id} - Q)(x_\nu - w_*) \\
&= -U(1 - \alpha) \sum_{\nu=1}^{n+1} U(x_\nu - w_*) \\
&= -Uv_{n+1} . \tag{22}
\end{aligned}$$

Under the previous considerations, (19) becomes

$$(1 - 2\alpha)(T - \text{Id})x_{n+1} + Q_\alpha(x_{n+1} - x_n) + Uv_{n+1} = -\lambda[\nabla f(x_n) + \xi_{n+1}] . \tag{23}$$

Recall now Proposition 15, and consider *any*  $(x_*, v_*) \in \mathcal{O}_\lambda$ . By the definition of  $\mathcal{O}_\lambda$ ,  $(\text{Id} - T)x_* = 0$  and there exists  $\xi_* \in \partial g(x_*)$  s.t.  $Uv_* + \lambda[\nabla f(x_*) + \xi_*] = 0$ . These arguments together with (23) yield

$$\lambda[\nabla f(x_n) - \nabla f(x_*)] + \lambda(\xi_{n+1} - \xi_*)$$

$$= -(1 - 2\alpha)(Q - \text{Id})(x_{n+1} - x_*) - Q_\alpha(x_{n+1} - x_n) - U(v_{n+1} - v_*). \quad (24)$$

The Baillon-Haddad theorem [1], [2, Cor. 18.16, p. 270] states that the  $L$ -Lipschitz continuous  $\nabla f$  is  $(1/L)$ -inverse strongly monotone, *i.e.*,  $\forall (x, x') \in \mathcal{X}^2$ ,  $\langle x - x' \mid \nabla f(x) - \nabla f(x') \rangle \geq (1/L) \|\nabla f(x) - \nabla f(x')\|^2$ . This property and the fact that  $\partial g$  is monotone [2, Example 20.3, p. 294], *i.e.*,  $\forall x, x', \xi, \xi'$  s.t.  $\xi \in \partial g(x)$  and  $\xi' \in \partial g(x')$ ,  $\langle x - x' \mid \xi - \xi' \rangle \geq 0$ , imply

$$\begin{aligned} & \frac{2\lambda}{L} \|\nabla f(x_n) - \nabla f(x_*)\|^2 \\ & \leq 2\lambda \langle x_n - x_* \mid \nabla f(x_n) - \nabla f(x_*) \rangle \\ & \leq 2\lambda \langle x_{n+1} - x_* \mid \nabla f(x_n) - \nabla f(x_*) \rangle + 2\lambda \langle x_n - x_{n+1} \mid \nabla f(x_n) - \nabla f(x_*) \rangle \\ & \quad + 2\lambda \langle x_{n+1} - x_* \mid \xi_{n+1} - \xi_* \rangle \\ & = 2\langle x_{n+1} - x_* \mid \lambda[\nabla f(x_n) - \nabla f(x_*)] + \lambda(\xi_{n+1} - \xi_*) \rangle \\ & \quad + 2\lambda \langle x_n - x_{n+1} \mid \nabla f(x_n) - \nabla f(x_*) \rangle \\ & = -2(1 - 2\alpha) \langle x_{n+1} - x_* \mid (Q - \text{Id})(x_{n+1} - x_*) \rangle - 2\langle x_{n+1} - x_* \mid Q_\alpha(x_{n+1} - x_n) \rangle \\ & \quad - 2\langle x_{n+1} - x_* \mid U(v_{n+1} - v_*) \rangle + 2\lambda \langle x_n - x_{n+1} \mid \nabla f(x_n) - \nabla f(x_*) \rangle \quad (25a) \\ & = -2(1 - 2\alpha) \langle x_{n+1} - x_* \mid (Q - \text{Id})(x_{n+1} - x_*) \rangle - 2\langle x_{n+1} - x_* \mid Q_\alpha(x_{n+1} - x_n) \rangle \\ & \quad - 2\langle U(x_{n+1} - x_*) \mid v_{n+1} - v_* \rangle + 2\lambda \langle x_n - x_{n+1} \mid \nabla f(x_n) - \nabla f(x_*) \rangle \\ & \leq -2(1 - 2\alpha) \langle x_{n+1} - x_* \mid (Q - \text{Id})(x_{n+1} - x_*) \rangle - 2\langle x_{n+1} - x_* \mid Q_\alpha(x_{n+1} - x_n) \rangle \\ & \quad - \frac{2}{1-\alpha} \langle v_{n+1} - v_n \mid v_{n+1} - v_* \rangle + \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2 \\ & \quad + \frac{2\lambda}{L} \|\nabla f(x_n) - \nabla f(x_*)\|^2, \quad (25b) \end{aligned}$$

where (24) was used in (25a), and (21) as well as

$$2 \left\langle \frac{a}{\sqrt{\eta}} \mid \sqrt{\eta} b \right\rangle_\Pi \leq \frac{1}{\eta} \|a\|_\Pi^2 + \eta \|b\|_\Pi^2, \quad \begin{cases} \forall (a, b) \in \mathcal{X}^2, \forall \eta \in \mathbb{R}_{>0}, \\ \forall \text{ strongly positive } \Pi \in \mathfrak{B}(\mathcal{X}), \end{cases} \quad (26)$$

with  $\eta := 2/L$ ,  $a := x_n - x_{n+1}$ ,  $b := \nabla f(x_n) - \nabla f(x_*)$ ,  $\Pi := \text{Id}$ , were used in (25b).

Recall (12) to verify that the positiveness of  $Q$  implies that for any  $x \in \mathcal{X}$ ,

$$\langle Q_\alpha x \mid x \rangle = \alpha \langle Qx \mid x \rangle + (1 - \alpha) \|x\|^2 \geq (1 - \alpha) \|x\|^2, \quad (27)$$

*i.e.*,  $Q_\alpha$  is strongly positive. Hence, upon defining the linear mapping  $\Theta : \mathcal{X}^2 \rightarrow \mathcal{X}^2 : (x, v) \mapsto (Q_\alpha x, v/(1 - \alpha))$ , it can be easily seen that  $\Theta$  is strongly positive, under the standard inner product  $\langle (x, v) \mid (x', v') \rangle := \langle x \mid x' \rangle + \langle v \mid v' \rangle$ ,  $\forall (x, v), (x', v') \in \mathcal{X}^2$ , due to the fact that both  $Q_\alpha$  and  $\text{Id}/(1 - \alpha)$  are strongly positive. Consequently,  $(\mathcal{X}^2, \langle \cdot \mid \cdot \rangle_\Theta)$  can be considered to be a Hilbert space equipped with the inner product  $\langle \cdot \mid \cdot \rangle_\Theta$ .

Notation  $y := (x, v)$ ,  $\alpha \geq 1/2$  as well as the positiveness of  $\text{Id} - Q$  in (25) yield

$$\begin{aligned} 0 & \leq 2 \langle (x_{n+1} - x_n, v_{n+1} - v_n) \mid \Theta(x_* - x_{n+1}, v_* - v_{n+1}) \rangle \\ & \quad - 2(2\alpha - 1) \langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle \\ & \quad + \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2 \end{aligned}$$

$$\begin{aligned}
&= 2\langle y_{n+1} - y_n \mid \Theta(y_* - y_{n+1}) \rangle - 2(2\alpha - 1)\langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle \\
&\quad + \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2 \\
&\leq 2\langle y_{n+1} - y_n \mid y_* - y_{n+1} \rangle_\Theta + \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2 \\
&= \|y_n - y_*\|_\Theta^2 - \|y_{n+1} - y_*\|_\Theta^2 - \|y_{n+1} - y_n\|_\Theta^2 + \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2.
\end{aligned}$$

Hence,

$$\|y_n - y_*\|_\Theta^2 - \|y_{n+1} - y_*\|_\Theta^2 \geq \|y_{n+1} - y_n\|_\Theta^2 - \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2. \quad (28)$$

Since  $\lambda < 2(1 - \alpha)/L$ , choose any  $\zeta \in (\lambda L/[2(1 - \alpha)], 1)$ . Then, by (27),  $\forall y := (x, v)$ ,

$$\frac{\lambda L}{2} \|x\|^2 < \zeta(1 - \alpha) \|x\|^2 \leq \zeta \langle x \mid Q_\alpha x \rangle \leq \zeta \langle x \mid Q_\alpha x \rangle + \zeta \frac{1}{1 - \alpha} \|v\|^2 = \zeta \|y\|_\Theta^2,$$

and by (28),

$$\begin{aligned}
\|y_n - y_*\|_\Theta^2 - \|y_{n+1} - y_*\|_\Theta^2 &\geq \|y_{n+1} - y_n\|_\Theta^2 - \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2 \\
&\geq \|y_{n+1} - y_n\|_\Theta^2 - \zeta \|y_{n+1} - y_n\|_\Theta^2 \\
&= (1 - \zeta) \|y_{n+1} - y_n\|_\Theta^2,
\end{aligned} \quad (29)$$

i.e., sequence  $(y_n)_{n \geq 2} \subset (\mathcal{X}^2, \langle \cdot \mid \cdot \rangle_\Theta)$  is Fejér monotone w.r.t.  $\mathcal{O}_\lambda$  of Proposition 15.

(ii) Due to Fejér monotonicity, sequence  $(y_n)_n$  is bounded [as well as  $(x_n)_n$  and  $(v_n)_n$ ] [2, Prop. 5.4(i), p. 76] and possesses a non-empty set of weakly sequential cluster points  $\mathfrak{W}[(y_n)_n]$  [2, Lem. 2.37, p. 36]. Moreover, it can be verified by (29), that  $\forall n \geq 2$ ,

$$(1 - \zeta) \sum_{\nu=2}^n \|y_{\nu+1} - y_\nu\|_\Theta^2 \leq \|y_2 - y_*\|_\Theta^2 - \|y_{n+1} - y_*\|_\Theta^2 \leq \|y_2 - y_*\|_\Theta^2,$$

and hence there exist  $C', C \in \mathbb{R}_{>0}$  s.t. for any  $n$ ,

$$\sum_{\nu=0}^n \|y_{\nu+1} - y_\nu\|_\Theta^2 \leq \frac{C'}{1 - \zeta} =: C, \quad (30)$$

which leads to  $\lim_{n \rightarrow \infty} \|y_{n+1} - y_n\|_\Theta = 0$ , and which further implies that

$$\lim_{n \rightarrow \infty} (x_{n+1} - x_n) = 0, \quad \lim_{n \rightarrow \infty} (v_{n+1} - v_n) = 0. \quad (31)$$

Adding the following equations, which result from (23),

$$\begin{aligned}
-\frac{1}{\lambda}(1 - 2\alpha)(T - \text{Id})x_{n+1} - \frac{1}{\lambda}Q_\alpha(x_{n+1} - x_n) - \frac{1}{\lambda}Uv_{n+1} - \nabla f(x_n) &= \xi_{n+1} \\
\frac{1}{\lambda}(1 - 2\alpha)(T - \text{Id})x_n + \frac{1}{\lambda}Q_\alpha(x_n - x_{n-1}) + \frac{1}{\lambda}Uv_n + \nabla f(x_{n-1}) &= -\xi_n
\end{aligned} \quad (32)$$

yields

$$\begin{aligned}
\xi_{n+1} - \xi_n &= \frac{1-2\alpha}{\lambda}(T - \text{Id})(x_n - x_{n+1}) + \frac{1}{\lambda}Q_\alpha(x_n - x_{n-1}) - \frac{1}{\lambda}Q_\alpha(x_{n+1} - x_n) \\
&\quad + \frac{1}{\lambda}U(v_n - v_{n+1}) + [\nabla f(x_{n-1}) - \nabla f(x_n)].
\end{aligned} \quad (33)$$

By applying  $\lim_{n \rightarrow \infty}$  to the previous equality, and by using the Lipschitz continuity of  $\nabla f$ , i.e.,  $\|\nabla f(x_n) - \nabla f(x_{n-1})\| \leq L\|x_n - x_{n-1}\|$ , (31), as well as the continuity of  $\text{Id} - T$ ,  $Q_\alpha$  and  $U$ , it can be verified that

$$\lim_{n \rightarrow \infty} (\xi_{n+1} - \xi_n) = 0. \quad (34)$$

Now, by (18),

$$\begin{aligned}
x_{n+2} - x_{n+1} &= Tx_{n+1} - T_\alpha x_{n+1} + T_\alpha x_{n+1} - T_\alpha x_n - \lambda[\nabla f(x_{n+1}) - \nabla f(x_n)] - \lambda[\xi_{n+2} - \xi_{n+1}] \\
&= (T - T_\alpha)x_{n+1} + Q_\alpha(x_{n+1} - x_n) - \lambda[\nabla f(x_{n+1}) - \nabla f(x_n)] - \lambda[\xi_{n+2} - \xi_{n+1}],
\end{aligned}$$

which leads to

$$\begin{aligned}
(1 - \alpha)(\text{Id} - T)x_n &= (x_n - x_{n+1}) + Q_\alpha(x_n - x_{n-1}) \\
&\quad - \lambda[\nabla f(x_n) - \nabla f(x_{n-1})] - \lambda[\xi_{n+1} - \xi_n].
\end{aligned} \tag{35}$$

Choose any  $\bar{y} := (\bar{x}, \bar{v}) \in \mathfrak{W}[(y_n)_{n \geq 2}] \neq \emptyset$ , i.e., there exists a subsequence  $(y_{n_k} := (x_{n_k}, v_{n_k}))_k$  s.t.  $x_{n_k} \rightharpoonup_{k \rightarrow \infty} \bar{x}$  and  $v_{n_k} \rightharpoonup_{k \rightarrow \infty} \bar{v}$ . Furthermore, by (31), (34), (35), and the Lipschitz continuity of  $\nabla f$ ,

$$\begin{aligned}
\limsup_{n \rightarrow \infty} \|(\text{Id} - T)x_n\| &\leq \frac{1}{1-\alpha} \lim_{k \rightarrow \infty} \|x_n - x_{n+1}\| + \lim_{k \rightarrow \infty} \frac{1}{1-\alpha} \|Q_\alpha(x_n - x_{n-1})\| \\
&\quad + \frac{\lambda}{1-\alpha} \lim_{k \rightarrow \infty} \|\nabla f(x_n) - \nabla f(x_{n-1})\| \\
&\quad + \frac{\lambda}{1-\alpha} \lim_{k \rightarrow \infty} \|\xi_{n+1} - \xi_n\| \\
&\leq \frac{1}{1-\alpha} \lim_{k \rightarrow \infty} \|x_n - x_{n+1}\| + \lim_{k \rightarrow \infty} \frac{\|Q_\alpha\|}{1-\alpha} \|x_n - x_{n-1}\| \\
&\quad + \frac{\lambda L}{1-\alpha} \lim_{k \rightarrow \infty} \|x_n - x_{n-1}\| + \frac{\lambda}{1-\alpha} \lim_{k \rightarrow \infty} \|\xi_{n+1} - \xi_n\| \\
&= 0.
\end{aligned} \tag{36}$$

Hence, due to  $x_{n_k} \rightharpoonup_{k \rightarrow \infty} \bar{x}$ ,  $\lim_{k \rightarrow \infty} (\text{Id} - T)x_{n_k} = 0$ , and the demiclosedness property of the nonexpansive mapping  $T$  [2, Thm. 4.17, p. 63], it follows that

$$\bar{x} \in \text{Fix } T. \tag{37}$$

Fix arbitrarily an  $x_\# \in \mathcal{X}$ . Since  $(x_n)_n$  is bounded, there exist  $C''$ ,  $C_{\nabla f} \in \mathbb{R}_{>0}$  s.t. for any  $n$ ,

$$\begin{aligned}
\|\nabla f(x_n)\| &\leq \|\nabla f(x_n) - \nabla f(x_\#)\| + \|\nabla f(x_\#)\| \\
&\leq L\|x_n - x_\#\| + \|\nabla f(x_\#)\| \\
&\leq L(\|x_n\| + \|x_\#\|) + \|\nabla f(x_\#)\| \\
&\leq L(C'' + \|x_\#\|) + \|\nabla f(x_\#)\| \leq C_{\nabla f}.
\end{aligned} \tag{38}$$

Now, according to the Baillon-Haddad theorem [1], [2, Cor. 18.16, p. 270],

$$\begin{aligned}
&\frac{2\lambda}{L} \|\nabla f(x_{n_k}) - \nabla f(\bar{x})\|^2 \\
&\leq 2\lambda \langle x_{n_k} - \bar{x} | \nabla f(x_{n_k}) - \nabla f(\bar{x}) \rangle \\
&= 2\lambda \langle x_{n_k+1} - \bar{x} | \nabla f(x_{n_k}) \rangle \\
&\quad - 2\lambda \langle x_{n_k+1} - \bar{x} | \nabla f(\bar{x}) \rangle + 2\lambda \langle x_{n_k} - x_{n_k+1} | \nabla f(x_{n_k}) - \nabla f(\bar{x}) \rangle \\
&= -2\lambda \langle x_{n_k+1} - \bar{x} | \xi_{n_k+1} \rangle - 2\langle x_{n_k+1} - \bar{x} | Uv_{n_k+1} \rangle
\end{aligned}$$

$$\begin{aligned}
& -2\langle x_{n_k+1} - \bar{x} \mid Q_\alpha(x_{n_k+1} - x_{n_k}) \rangle - (1 - 2\alpha)\langle x_{n_k+1} - \bar{x} \mid (T - \text{Id})x_{n_k+1} \rangle \\
& -2\lambda\langle x_{n_k+1} - \bar{x} \mid \nabla f(\bar{x}) \rangle + 2\lambda\langle x_{n_k} - x_{n_k+1} \mid \nabla f(x_{n_k}) - \nabla f(\bar{x}) \rangle
\end{aligned} \tag{39a}$$

$$\begin{aligned}
& \leq 2\lambda[g(\bar{x}) - g(x_{n_k+1})] - 2\langle U(x_{n_k+1} - \bar{x}) \mid v_{n_k+1} \rangle \\
& -2\langle x_{n_k+1} - \bar{x} \mid Q_\alpha(x_{n_k+1} - x_{n_k}) \rangle - (1 - 2\alpha)\langle x_{n_k+1} - \bar{x} \mid (T - \text{Id})x_{n_k+1} \rangle \\
& -2\lambda\langle x_{n_k+1} - \bar{x} \mid \nabla f(\bar{x}) \rangle + 2\lambda\langle x_{n_k} - x_{n_k+1} \mid \nabla f(x_{n_k}) - \nabla f(\bar{x}) \rangle
\end{aligned} \tag{39b}$$

$$\begin{aligned}
& \leq 2\lambda[g(\bar{x}) - g(x_{n_k+1})] - \frac{2}{1-\alpha}\langle v_{n_k+1} - v_{n_k} \mid v_{n_k+1} \rangle \\
& -2\langle x_{n_k+1} - \bar{x} \mid Q_\alpha(x_{n_k+1} - x_{n_k}) \rangle - (1 - 2\alpha)\langle x_{n_k+1} - \bar{x} \mid (T - \text{Id})x_{n_k+1} \rangle \\
& -2\lambda\langle x_{n_k+1} - \bar{x} \mid \nabla f(\bar{x}) \rangle + 2\lambda(C_{\nabla f} + \|\nabla f(\bar{x})\|) \|x_{n_k} - x_{n_k+1}\|,
\end{aligned} \tag{39c}$$

where (23) was used in (39a), the convexity of  $g$ , (17) and the self-adjointness of  $U$  in (39b), and finally (21) and (38) in (39c). Since  $\lim_{k \rightarrow \infty} (x_{n_k} - x_{n_k+1}) = 0$  by (31), the continuity of  $Q_\alpha$  implies  $\lim_{k \rightarrow \infty} Q_\alpha(x_{n_k+1} - x_{n_k}) = 0$ , and (36) yields  $\lim_{k \rightarrow \infty} (T - \text{Id})x_{n_k+1} = 0$ . Notice again by (31) that  $\lim_{k \rightarrow \infty} (v_{n_k+1} - v_{n_k}) = 0$ . Further, (31), together with  $(x_{n_k} - \bar{x}) \rightarrow_{k \rightarrow \infty} 0$ , yields  $(x_{n_k+1} - \bar{x}) \rightarrow_{k \rightarrow \infty} 0$ . Similarly,  $(v_{n_k+1} - \bar{v}) \rightarrow_{k \rightarrow \infty} 0$  can be deduced from (31) and  $(v_{n_k} - \bar{v}) \rightarrow_{k \rightarrow \infty} 0$ . Due to [2, Lem. 2.41(iii), p. 37], all of the previous arguments result in  $\lim_{k \rightarrow \infty} \langle v_{n_k+1} - v_{n_k} \mid v_{n_k+1} \rangle = 0$ ,  $\lim_{k \rightarrow \infty} \langle x_{n_k+1} - \bar{x} \mid Q_\alpha(x_{n_k+1} - x_{n_k}) \rangle = 0$ ,  $\lim_{k \rightarrow \infty} \langle x_{n_k+1} - \bar{x} \mid (T - \text{Id})x_{n_k+1} \rangle = 0$ ,  $\lim_{k \rightarrow \infty} \langle x_{n_k+1} - \bar{x} \mid \nabla f(\bar{x}) \rangle = 0$ , and  $\lim_{k \rightarrow \infty} \|x_{n_k} - x_{n_k+1}\| = 0$ . Hence, the application of  $\limsup_{k \rightarrow \infty}$  onto both sides of (39c) yields

$$\begin{aligned}
\limsup_{k \rightarrow \infty} \|\nabla f(x_{n_k}) - \nabla f(\bar{x})\|^2 & \leq \limsup_{k \rightarrow \infty} L[g(\bar{x}) - g(x_{n_k+1})] \\
& = L \left[ g(\bar{x}) - \liminf_{k \rightarrow \infty} g(x_{n_k+1}) \right] \leq 0,
\end{aligned}$$

where the last inequality is deduced from the fact that  $g \in \Gamma_0(\mathcal{X})$  turns out to be also weakly sequentially lower semicontinuous [2, Thm. 9.1, p. 129]. In other words,

$$\lim_{k \rightarrow \infty} \nabla f(x_{n_k}) = \nabla f(\bar{x}). \tag{40}$$

Since  $v_{n_k+1} \rightarrow_k \bar{v}$ , i.e.,  $\forall z \in \mathcal{X}$ ,  $\lim_{k \rightarrow \infty} \langle z \mid v_{n_k+1} \rangle = \langle z \mid \bar{v} \rangle$ , it can be easily seen that  $\forall z \in \mathcal{X}$ ,  $\lim_{k \rightarrow \infty} \langle z \mid Uv_{n_k+1} \rangle = \lim_{k \rightarrow \infty} \langle Uz \mid v_{n_k+1} \rangle = \langle Uz \mid \bar{v} \rangle = \langle z \mid U\bar{v} \rangle$ , i.e.,  $Uv_{n_k+1} \rightarrow_k U\bar{v}$ . Hence, having this result and (40) plugged into (32) yields that

$$\xi_{n_k+1} \rightarrow_{k \rightarrow \infty} \bar{\xi} := -\frac{1}{\lambda}U\bar{v} - \nabla f(\bar{x}). \tag{41}$$

Using (23) once again,

$$\begin{aligned}
\langle x_{n_k+1} - \bar{x} \mid \xi_{n_k+1} \rangle & = -\langle x_{n_k+1} - \bar{x} \mid \nabla f(x_{n_k}) \rangle - \frac{1}{\lambda}\langle x_{n_k+1} - \bar{x} \mid Uv_{n_k+1} \rangle \\
& \quad - \frac{1}{\lambda}\langle x_{n_k+1} - \bar{x} \mid Q_\alpha(x_{n_k+1} - x_{n_k}) \rangle \\
& \quad - \frac{1}{\lambda}(1 - 2\alpha)\langle x_{n_k+1} - \bar{x} \mid (T - \text{Id})x_{n_k+1} \rangle \\
& = -\langle x_{n_k+1} - \bar{x} \mid \nabla f(x_{n_k}) \rangle - \frac{1}{\lambda}\langle U(x_{n_k+1} - \bar{x}) \mid v_{n_k+1} \rangle \\
& \quad - \frac{1}{\lambda}\langle x_{n_k+1} - \bar{x} \mid Q_\alpha(x_{n_k+1} - x_{n_k}) \rangle
\end{aligned}$$



$$\begin{aligned}
& -\frac{1}{\lambda}(1-2\alpha)\langle x_{n_k+1} - \bar{x} \mid (T - \text{Id})x_{n_k+1} \rangle \\
& = -\langle x_{n_k+1} - \bar{x} \mid \nabla f(x_{n_k}) \rangle - \frac{1}{\lambda(1-\alpha)}\langle v_{n_k+1} - v_{n_k} \mid v_{n_k+1} \rangle \\
& -\frac{1}{\lambda}\langle x_{n_k+1} - \bar{x} \mid Q_\alpha(x_{n_k+1} - x_{n_k}) \rangle \\
& -\frac{1}{\lambda}(1-2\alpha)\langle x_{n_k+1} - \bar{x} \mid (T - \text{Id})x_{n_k+1} \rangle, \tag{42}
\end{aligned}$$

where (21) was used in (42). Since  $(x_{n_k+1} - \bar{x}) \rightharpoonup_k 0$  and  $v_{n_k+1} \rightharpoonup_k \bar{v}$ , and due to (31), (36) and (40), as well as the continuity of the linear mapping  $Q_\alpha$ , it turns out by [2, Lem. 2.41(iii), p. 37] and (42) that  $\lim_{k \rightarrow \infty} \langle x_{n_k+1} - \bar{x} \mid \xi_{n_k+1} \rangle = 0$ . In other words,

$$\begin{aligned}
\lim_{k \rightarrow \infty} \langle x_{n_k+1} \mid \xi_{n_k+1} \rangle & = \lim_{k \rightarrow \infty} (\langle x_{n_k+1} - \bar{x} \mid \xi_{n_k+1} \rangle + \langle \bar{x} \mid \xi_{n_k+1} \rangle) \\
& = \lim_{k \rightarrow \infty} \langle x_{n_k+1} - \bar{x} \mid \xi_{n_k+1} \rangle + \lim_{k \rightarrow \infty} \langle \bar{x} \mid \xi_{n_k+1} \rangle \\
& = \lim_{k \rightarrow \infty} \langle \bar{x} \mid \xi_{n_k+1} \rangle = \langle \bar{x} \mid \bar{\xi} \rangle. \tag{43}
\end{aligned}$$

Now, by  $(x_{n_k+1}, \xi_{n_k+1}) \in \text{gra } \partial g$ , the maximal monotonicity of  $\partial g$  [2, Thm. 20.40, p. 304] and the property manifested in (43), [2, Cor. 20.49(ii), p. 306] suggests that  $(\bar{x}, \bar{\xi}) \in \text{gra } \partial g \Leftrightarrow \bar{\xi} \in \partial g(\bar{x})$ . Hence, according also to (41),  $-U(\bar{v}/\lambda) \in \nabla f(\bar{x}) + \partial g(\bar{x})$ , which together with (37) imply  $(\bar{x}, \bar{v}) \in \mathcal{O}_\lambda$ . Since  $(\bar{x}, \bar{v})$  was arbitrarily chosen within  $\mathfrak{W}[(y_n)_n]$ , it follows that  $\mathfrak{W}[(y_n)_n] \subset \mathcal{O}_\lambda$ . Adding also to that the Fejér monotonicity property (29) of  $(y_n)_{n \geq 2}$  w.r.t.  $\mathcal{O}_\lambda$  yields that  $(y_n)_n$  converges weakly to a point in  $\mathcal{O}_\lambda$  [2, Thm. 5.5, p. 76]. According to (9c), the weak limit of  $(x_n)_n$  solves  $\text{VIP}(\nabla f + \partial g, \text{Fix } T)$ .

(iii) As part (ii) of the proof has demonstrated, sequences  $(x_n)_n$  and  $(Uv_n)_n$  converge weakly to  $\bar{x}$  and  $U\bar{v}$ , respectively. Consequently, (31), the continuity of  $Q_\alpha$ , (32), (36), (40) and (41) suggest that  $(\xi_n)_n$  converges weakly to  $\bar{\xi}$ .

Let Assumption 17(i) hold true. Then, according to [2, Ex. 22.3(iii), p. 324], given a bounded set  $\mathcal{B} \subset \mathcal{X}$ , there exists an increasing function  $\varphi_{\mathcal{B}} : [0, +\infty) \rightarrow [0, +\infty]$ , which vanishes only at 0, s.t.  $\forall x, x' \in \mathcal{B}$ ,

$$\langle x - x' \mid \nabla f(x) - \nabla f(x') \rangle \geq 2\varphi_{\mathcal{B}}(\|x - x'\|). \tag{44}$$

Define  $\mathcal{B} := (x_n)_n \cup \{\bar{x}\}$  (recall that  $(x_n)_n$  is bounded). Set  $x := x_n$  and  $x' := \bar{x}$  in (44) to obtain

$$\langle x_n - \bar{x} \mid \nabla f(x_n) - \nabla f(\bar{x}) \rangle \geq 2\varphi_{\mathcal{B}}(\|x_n - \bar{x}\|), \quad \forall n. \tag{45}$$

Since  $x_n \rightharpoonup_{n \rightarrow \infty} \bar{x}$  and  $\lim_{n \rightarrow \infty} \nabla f(x_n) = \nabla f(\bar{x})$  by (40), the application of  $\lim_{n \rightarrow \infty}$  to (45) and [2, Lem. 2.41(iii), p. 37] suggest that  $\lim_{n \rightarrow \infty} \varphi_{\mathcal{B}}(\|x_n - \bar{x}\|) = 0$ , and thus  $\lim_{n \rightarrow \infty} \|x_n - \bar{x}\| = 0$ , due to the properties of  $\varphi_{\mathcal{B}}$ .

Let now Assumption 17(ii) hold true. Then, according to [2, Ex. 22.3(iii), p. 324], given a bounded set  $\mathcal{B} \subset \text{dom } \partial g$ , there exists an increasing function  $\varphi_{\mathcal{B}} : [0, +\infty) \rightarrow [0, +\infty]$ , which vanishes only at 0, s.t.  $\forall x, x' \in \mathcal{B}$ , and  $\forall \xi \in \partial g(x)$ ,  $\forall \xi' \in \partial g(x')$ ,

$$\langle x - x' \mid \xi - \xi' \rangle \geq 2\varphi_{\mathcal{B}}(\|x - x'\|). \tag{46}$$

According to (17),  $x_n \in \text{dom } \partial g$ ,  $\forall n$ . Moreover, as the discussion after (43) demonstrated,  $\bar{x} \in \text{dom } \partial g$ . Define thus the bounded set  $\mathcal{B} := (x_n)_n \cup \{\bar{x}\} \subset \text{dom } \partial g$ , and set  $x := x_n$ ,  $x' := \bar{x}$ ,  $\xi := \xi_n$  and  $\xi' := \bar{\xi}$  in (46) to obtain

$$\langle x_n - \bar{x} \mid \xi_n - \bar{\xi} \rangle \geq 2\varphi_{\mathcal{B}}(\|x_n - \bar{x}\|), \quad \forall n. \quad (47)$$

Similarly to (43), it can be verified that  $\lim_{n \rightarrow \infty} \langle x_n \mid \xi_n \rangle = \langle \bar{x} \mid \bar{\xi} \rangle$ . Thus,

$$\begin{aligned} \lim_{n \rightarrow \infty} \langle x_n - \bar{x} \mid \xi_n - \bar{\xi} \rangle &= \lim_{n \rightarrow \infty} \langle x_n \mid \xi_n \rangle - \lim_{n \rightarrow \infty} \langle x_n \mid \bar{\xi} \rangle - \lim_{n \rightarrow \infty} \langle \bar{x} \mid \xi_n \rangle + \langle \bar{x} \mid \bar{\xi} \rangle \\ &= \langle \bar{x} \mid \bar{\xi} \rangle - \langle \bar{x} \mid \bar{\xi} \rangle - \langle \bar{x} \mid \bar{\xi} \rangle + \langle \bar{x} \mid \bar{\xi} \rangle = 0. \end{aligned}$$

Hence, the application of  $\lim_{n \rightarrow \infty}$  to (47) yields  $\lim_{n \rightarrow \infty} \varphi_{\mathcal{B}}(\|x_n - \bar{x}\|) = 0$ , and thus  $\lim_{n \rightarrow \infty} \|x_n - \bar{x}\| = 0$ .  $\square$

The following theorem draws even stronger links with the original form of HSDM.

**Theorem 19.** Consider  $f \in \Gamma_0(\mathcal{X})$ , with  $L$  being the Lipschitz-continuity constant of  $\nabla f$ . Moreover, given the closed affine set  $\mathcal{A}$ , consider any  $T \in \mathfrak{T}_{\mathcal{A}}$ , and for  $\lambda \in \mathbb{R}_{>0}$ , an arbitrarily fixed  $x_0 \in \mathcal{X}$ , and for all  $n \in \mathbb{Z}_{\geq 0}$  form the iterations:

$$x_{1/2} := T_{\alpha}x_0 - \lambda \nabla f(T_{\alpha}x_0), \quad (48a)$$

$$x_1 := x_{1/2}, \quad (48b)$$

$$x_{n+3/2} := x_{n+1/2} - [T_{\alpha}x_n - \lambda \nabla f(T_{\alpha}x_n)] + [Tx_{n+1} - \lambda \nabla f(T_{\alpha}x_{n+1})], \quad (48c)$$

$$x_{n+2} := x_{n+3/2}, \quad (48d)$$

where  $T_{\alpha}$  is defined in (12).

Consider  $\alpha \in [0.5, 1)$  and  $\lambda \in (0, 2(1 - \alpha)^2/L)$ . Then, the following hold true.

- (i) There exist a sequence  $(v_n)_{n \in \mathbb{Z}_{\geq 0}} \subset \mathcal{X}$  and a strongly positive operator  $\Upsilon : \mathcal{X}^2 \rightarrow \mathcal{X}^2$  s.t. sequence  $(y_n := (x_n, v_n))_{n \in \mathbb{Z}_{>0} \setminus \{1\}}$  is Fejér monotone [2, Def. 5.1, p. 75] w.r.t.  $\mathcal{O}_{\lambda}$  of Proposition 15 (under  $g = 0$ ) in the Hilbert space  $(\mathcal{X}^2, \langle \cdot \mid \cdot \rangle_{\Upsilon})$ .
- (ii) Sequence  $(x_n)_n$  of (48) converges weakly to a point that solves  $\text{VIP}(\nabla f, \text{Fix } T)$ .
- (iii) If Assumption 17(i) also holds true, then  $(x_n)_n$  of (48) converges strongly to a point that solves  $\text{VIP}(\nabla f, \text{Fix } T)$ .  $\square$

*Proof.* (i) Proposition 15 takes the following special form in the present context: If  $\exists v_* \in \mathcal{X}$  s.t.

$$(x_*, v_*) \in \mathcal{O}_{\lambda} := \{(x, v) \in \text{Fix } T \times \mathcal{X} \mid -\frac{1}{\lambda}Uv = \nabla f(x)\}, \quad (49)$$

then  $x_*$  solves  $\text{VIP}(\nabla f, \text{Fix } T)$ .

By following the same steps which start from the beginning of the proof of Theorem 18 till (22), it can be verified that

$$-(1 - 2\alpha)(T - \text{Id})x_{n+1} - Q_{\alpha}(x_{n+1} - x_n) - Uv_{n+1} = \lambda \nabla f(T_{\alpha}x_n), \quad (50)$$

and by considering any  $(x_*, v_*) \in \mathcal{O}_\lambda$ ,

$$\begin{aligned} & \lambda[\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)] \\ &= -(1 - 2\alpha)(Q - \text{Id})(x_{n+1} - x_*) - Q_\alpha(x_{n+1} - x_n) - U(v_{n+1} - v_*). \end{aligned} \quad (51)$$

As in the proof of Theorem 18, the Baillon-Haddad theorem [1], [2, Cor. 18.16, p. 270] suggests that

$$\begin{aligned} & \frac{2\lambda}{L} \|\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)\|^2 \\ & \leq 2\lambda \langle T_\alpha x_n - T_\alpha x_* \mid \nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*) \rangle \\ & = 2\lambda \langle Q_\alpha(x_n - x_*) \mid \nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*) \rangle \\ & = 2\lambda \langle x_n - x_* \mid Q_\alpha[\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)] \rangle \\ & = 2\lambda \langle x_{n+1} - x_* \mid Q_\alpha[\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)] \rangle \\ & \quad + 2\lambda \langle x_n - x_{n+1} \mid Q_\alpha[\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)] \rangle \\ & = -2(1 - 2\alpha) \langle x_{n+1} - x_* \mid Q_\alpha(Q - \text{Id})(x_{n+1} - x_*) \rangle - 2 \langle x_{n+1} - x_* \mid Q_\alpha^2(x_{n+1} - x_n) \rangle \\ & \quad - 2 \langle x_{n+1} - x_* \mid Q_\alpha U(v_{n+1} - v_*) \rangle \\ & \quad + 2\lambda \langle x_n - x_{n+1} \mid Q_\alpha[\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)] \rangle \\ & = -2(1 - 2\alpha) \langle x_{n+1} - x_* \mid Q_\alpha(Q - \text{Id})(x_{n+1} - x_*) \rangle - 2 \langle x_{n+1} - x_* \mid Q_\alpha^2(x_{n+1} - x_n) \rangle \\ & \quad - 2 \langle U(x_{n+1} - x_*) \mid Q_\alpha(v_{n+1} - v_*) \rangle + 2\lambda \langle x_n - x_{n+1} \mid Q_\alpha[\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)] \rangle \\ & \leq -2(1 - 2\alpha) \langle x_{n+1} - x_* \mid Q_\alpha(Q - \text{Id})(x_{n+1} - x_*) \rangle - 2 \langle x_{n+1} - x_* \mid Q_\alpha^2(x_{n+1} - x_n) \rangle \\ & \quad - \frac{2}{1-\alpha} \langle v_{n+1} - v_n \mid Q_\alpha(v_{n+1} - v_*) \rangle + \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2 \\ & \quad + \frac{2\lambda}{L} \|Q_\alpha[\nabla f(x_n) - \nabla f(x_*)]\|^2 \\ & \leq -2(2\alpha - 1) \langle x_{n+1} - x_* \mid Q_\alpha(\text{Id} - Q)(x_{n+1} - x_*) \rangle - 2 \langle x_{n+1} - x_* \mid Q_\alpha^2(x_{n+1} - x_n) \rangle \\ & \quad - \frac{2}{1-\alpha} \langle v_{n+1} - v_n \mid Q_\alpha(v_{n+1} - v_*) \rangle + \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2 \\ & \quad + \frac{2\lambda}{L} \|\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)\|^2 \\ & \leq 2 \langle x_* - x_{n+1} \mid Q_\alpha^2(x_{n+1} - x_n) \rangle + \frac{2}{1-\alpha} \langle v_{n+1} - v_n \mid Q_\alpha(v_* - v_{n+1}) \rangle \\ & \quad + \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2 + \frac{2\lambda}{L} \|\nabla f(T_\alpha x_n) - \nabla f(T_\alpha x_*)\|^2. \end{aligned} \quad (52)$$

Mapping  $Q_\alpha^2$  is strongly positive: Indeed, if  $U_\alpha$  denotes the square root of the strongly positive  $Q_\alpha$  [cf. (27)], then  $\forall x \in \mathcal{X}$ ,  $\langle Q_\alpha^2 x \mid x \rangle = \langle U_\alpha Q_\alpha U_\alpha x \mid x \rangle = \langle Q_\alpha U_\alpha x \mid U_\alpha x \rangle \geq (1 - \alpha) \langle U_\alpha x \mid U_\alpha x \rangle = (1 - \alpha) \langle Q_\alpha x \mid x \rangle \geq (1 - \alpha)^2 \|x\|^2$ . Define now the mapping  $\Upsilon : \mathcal{X}^2 \rightarrow \mathcal{X}^2 : (x, v) \mapsto (Q_\alpha^2 x, [1/(1 - \alpha)]Q_\alpha v)$ . Mapping  $\Upsilon$  turns out to be strongly positive, w.r.t. the standard inner product of  $\mathcal{X}^2$ :  $\langle (x, v) \mid (x', v') \rangle := \langle x \mid x' \rangle + \langle v \mid v' \rangle$ ,  $\forall (x, v), (x', v') \in \mathcal{X}^2$ , due to the strong positiveness of  $Q_\alpha^2$  and  $[1/(1 - \alpha)]Q_\alpha$ . Consequently, one can consider  $(\mathcal{X}^2, \langle \cdot \mid \cdot \rangle_\Upsilon)$  as a Hilbert space equipped with the inner product  $\langle (x, v) \mid (x', v') \rangle_\Upsilon := \langle x \mid Q_\alpha^2 x' \rangle + [1/(1 - \alpha)] \langle v \mid Q_\alpha v' \rangle$ ,  $\forall (x, v), (x', v') \in \mathcal{X}^2$ . As such, (52) becomes

$$\begin{aligned} 0 & \leq 2 \langle y_{n+1} - y_n \mid \Upsilon(y_* - y_{n+1}) \rangle + \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2 \\ & = 2 \langle y_{n+1} - y_n \mid y_* - y_{n+1} \rangle_\Upsilon + \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2 \end{aligned}$$

$$= \|y_n - y_*\|_{\Upsilon}^2 - \|y_{n+1} - y_*\|_{\Upsilon}^2 - \|y_{n+1} - y_n\|_{\Upsilon}^2 + \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2. \quad (53)$$

Choose, now, any  $\zeta'$  with  $\lambda L/[2(1-\alpha)^2] < \zeta' < 1$ . Then, for any  $y = (x, v) \in \mathcal{X}^2$ ,

$$\begin{aligned} \frac{\lambda L}{2} \|x\|^2 &< \zeta'(1-\alpha)^2 \|x\|^2 \leq \zeta' \langle x \mid Q_{\alpha}^2 x \rangle \\ &\leq \zeta' \langle x \mid Q_{\alpha}^2 x \rangle + \zeta' \frac{1}{1-\alpha} \langle v \mid Q_{\alpha} v \rangle = \zeta' \|y\|_{\Upsilon}^2. \end{aligned}$$

This argument together with (53) yield

$$\begin{aligned} \|y_n - y_*\|_{\Upsilon}^2 - \|y_{n+1} - y_*\|_{\Upsilon}^2 &\geq \|y_{n+1} - y_n\|_{\Upsilon}^2 - \frac{\lambda L}{2} \|x_n - x_{n+1}\|^2 \\ &\geq \|y_{n+1} - y_n\|_{\Upsilon}^2 - \zeta' \|y_{n+1} - y_n\|_{\Upsilon}^2 \\ &= (1 - \zeta') \|y_{n+1} - y_n\|_{\Upsilon}^2, \end{aligned} \quad (54)$$

i.e., sequence  $(y_n)_{n \geq 2} \subset (\mathcal{X}^2, \langle \cdot \mid \cdot \rangle_{\Upsilon})$  is Fejér monotone w.r.t.  $\mathcal{O}_{\lambda}$  of (49).

(ii) Due to Fejér monotonicity,  $(y_n)$  is bounded [2, Prop. 5.4(i), p. 76] and possesses a non-empty set of weakly sequential cluster points  $\mathfrak{W}[(y_n)_n]$  [2, Lem. 2.37, p. 36]. Moreover, it can be readily verified, as in (31), that  $\lim_{n \rightarrow \infty} (y_{n+1} - y_n) = 0$ ,  $\lim_{n \rightarrow \infty} (x_{n+1} - x_n) = 0$  and  $\lim_{n \rightarrow \infty} (v_{n+1} - v_n) = 0$ . The rest of the proof follows steps similar to those after (31) in the proof of Theorem 18, but with the following twist:  $\nabla f(x_{n_k})$  is replaced by  $\nabla f(T_{\alpha} x_{n_k})$ , where all the asymptotic results of the proof of Theorem 18 continue to hold due to the Lipschitz continuity of  $\nabla f$  and the nonexpansiveness of  $T_{\alpha}$ , e.g.,  $\forall x, x' \in \mathcal{X}$ ,

$$\|\nabla f(T_{\alpha} x) - \nabla f(T_{\alpha} x')\| \leq L \|T_{\alpha} x - T_{\alpha} x'\| \leq L \|x - x'\|.$$

(iii) Part (ii) of this proof has demonstrated that sequences  $(x_n)_n$  and  $(Uv_n)_n$  converge weakly to  $\bar{x}$  and  $U\bar{v}$ , respectively. Consequently, in a way similar to part (ii) of the proof of Theorem 18, it can be shown also here that  $(\xi_n)_n$  converges weakly to  $\bar{\xi}$ .

Let Assumption 17(i) hold true. Then, according to [2, Ex. 22.3(iii), p. 324], given a bounded set  $\mathcal{B} \subset \mathcal{X}$ , there exists an increasing function  $\varphi_{\mathcal{B}} : [0, +\infty) \rightarrow [0, +\infty]$ , which vanishes only at 0, s.t.  $x, x' \in \mathcal{B}$ ,

$$\langle x - x' \mid \nabla f(x) - \nabla f(x') \rangle \geq 2\varphi_{\mathcal{B}}(\|x - x'\|). \quad (55)$$

Due to the nonexpansiveness of  $T_{\alpha}$  and the boundedness of  $(x_n)_n$ , by part (i) of the proof, it turns out that  $(T_{\alpha} x_n)_n$  is also bounded:  $\|T_{\alpha} x_n\| \leq \|T_{\alpha} x_n - T_{\alpha} \bar{x}\| + \|T_{\alpha} \bar{x}\| \leq \|x_n - \bar{x}\| + \|\bar{x}\| \leq \|x_n\| + 2\|\bar{x}\| \leq C'' + 2\|\bar{x}\|$ , for some  $C'' \in \mathbb{R}_{>0}$  (recall that  $\bar{x} \in \text{Fix } T_{\alpha} = \text{Fix } T$ ). Define, thus, the bounded set  $\mathcal{B} := (T_{\alpha} x_n)_n \cup \{\bar{x}\}$ . As such, (55) yields

$$\begin{aligned} &\langle (T_{\alpha} - \text{Id})x_n \mid \nabla f(T_{\alpha} x_n) - \nabla f(\bar{x}) \rangle + \langle x_n - \bar{x} \mid \nabla f(T_{\alpha} x_n) - \nabla f(\bar{x}) \rangle \\ &= \langle T_{\alpha} x_n - \bar{x} \mid \nabla f(T_{\alpha} x_n) - \nabla f(\bar{x}) \rangle \geq 2\varphi_{\mathcal{B}}(\|T_{\alpha} x_n - \bar{x}\|), \quad \forall n. \end{aligned} \quad (56)$$

Part (i) of this proof has already showed that  $\lim_{n \rightarrow \infty} (T - \text{Id})x_n = 0$ . As such,  $\lim_{n \rightarrow \infty} (T_{\alpha} - \text{Id})x_n = \alpha \lim_{n \rightarrow \infty} (T - \text{Id})x_n = 0$ . Moreover, note that  $x_n \rightharpoonup_{n \rightarrow \infty} \bar{x}$ ,

and  $\lim_{n \rightarrow \infty} \nabla f(T_\alpha x_n) = \nabla f(\bar{x})$ . Hence, due also to [2, Lem. 2.41(iii), p. 37], an application of  $\lim_{n \rightarrow \infty}$  to both sides of (56) results in  $\lim_{n \rightarrow \infty} \varphi_B(\|T_\alpha x_n - \bar{x}\|) = 0$ , and thus  $\lim_{n \rightarrow \infty} T_\alpha x_n = \bar{x}$ . Using  $\lim_{n \rightarrow \infty} (T_\alpha - \text{Id})x_n = 0$ , one can easily verify that  $\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} (\text{Id} - T_\alpha)x_n + \lim_{n \rightarrow \infty} T_\alpha x_n = \bar{x}$ , which establishes part (ii) of Theorem 19.  $\square$

The following theorems present convergence rates on the sequence of FM-HSDM estimates.

**Theorem 20.** For sequence  $(x_n)_{n \in \mathbb{Z}_{\geq 0}}$  of (13), there exists  $\xi_n \in \partial g(x_n)$ ,  $\forall n$ , s.t. for any  $x_* \in \text{Fix } T$ ,

$$\frac{1}{n+1} \sum_{\nu=0}^n \langle x_{\nu+1} - x_* \mid (\text{Id} - Q)(x_{\nu+1} - x_*) \rangle = O\left(\frac{1}{n+1}\right), \quad (57a)$$

$$\frac{1}{n+1} \sum_{\nu=0}^n \|Uv_{\nu+1} + \lambda[\nabla f(x_\nu) + \xi_{\nu+1}]\|^2 = O\left(\frac{1}{n+1}\right), \quad (57b)$$

$$\frac{1}{n+1} \sum_{\nu=0}^n \|(\text{Id} - T)x_{\nu+1}\|^2 = O\left(\frac{1}{n+1}\right), \quad (57c)$$

where the big-oh notation  $a_n = O(b_n)$ ,  $b_n > 0$ , means  $\limsup_{n \rightarrow \infty} |a_n|/b_n < +\infty$ . Regarding sequence  $(x_n)_{n \in \mathbb{Z}_{\geq 0}}$  of (14), (57a)–(57c) still hold true, but  $\xi_{\nu+1}$  is set equal to 0 in (57b). Similarly, for sequence  $(x_n)_{n \in \mathbb{Z}_{\geq 0}}$  of (48), (57a), (57c) as well as

$$\frac{1}{n+1} \sum_{\nu=0}^n \|Uv_{\nu+1} + \lambda \nabla f(T_\alpha x_\nu)\|^2 = O\left(\frac{1}{n+1}\right)$$

hold true.  $\square$

*Proof.* First, notice by (27), Proposition 26 and  $\|Q_\alpha\| \leq 1$  that  $Q_\alpha^{-1}$  exists and it is strongly positive with

$$\|Q_\alpha^{-1}\| \leq \frac{1}{1-\alpha}; \quad (1-\alpha)\|x\|^2 \leq \frac{(1-\alpha)}{\|Q_\alpha\|^2} \|x\|^2 \leq \langle Q_\alpha^{-1}x \mid x \rangle, \quad \forall x \in \mathcal{X}. \quad (58)$$

Then, going back to the discussion following (27),

$$\begin{aligned} & \|y_{n+1} - y_n\|_\Theta^2 \\ &= \|x_{n+1} - x_n\|_{Q_\alpha}^2 + \frac{1}{1-\alpha} \|v_{n+1} - v_n\|^2 \end{aligned} \quad (59a)$$

$$= \|Q_\alpha(x_{n+1} - x_n)\|_{Q_\alpha^{-1}}^2 + \frac{1}{1-\alpha} \|(1-\alpha)U(x_{n+1} - x_*)\|^2 \quad (59b)$$

$$\begin{aligned} &= \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}] - (1-2\alpha)(\text{Id} - T)x_{n+1}\|_{Q_\alpha^{-1}}^2 \\ &\quad + \frac{1}{1-\alpha} \|(1-\alpha)U(x_{n+1} - x_*)\|^2 \end{aligned} \quad (59c)$$

$$\begin{aligned} &= \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|_{Q_\alpha^{-1}}^2 + (1-2\alpha)^2 \|(\text{Id} - T)x_{n+1}\|_{Q_\alpha^{-1}}^2 \\ &\quad - 2\langle Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}] \mid (1-2\alpha)(\text{Id} - T)x_{n+1} \rangle_{Q_\alpha^{-1}} \\ &\quad + \frac{1}{1-\alpha} \|(1-\alpha)U(x_{n+1} - x_*)\|^2 \\ &\geq \frac{1}{\rho} \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|_{Q_\alpha^{-1}}^2 - \frac{(1-2\alpha)^2}{\rho-1} \|(\text{Id} - T)x_{n+1}\|_{Q_\alpha^{-1}}^2 \\ &\quad + \frac{1}{1-\alpha} \|(1-\alpha)U(x_{n+1} - x_*)\|^2 \end{aligned} \quad (59d)$$

$$= \frac{1}{\rho} \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|_{Q_\alpha^{-1}}^2 - \frac{(1-2\alpha)^2}{\rho-1} \|(\text{Id} - T)x_{n+1} - (\text{Id} - T)x_*\|_{Q_\alpha^{-1}}^2$$

$$\begin{aligned}
& + (1 - \alpha) \langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle \\
& = \frac{1}{\rho} \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|_{Q_\alpha^{-1}}^2 - \frac{(1-2\alpha)^2}{\rho-1} \|(\text{Id} - Q)(x_{n+1} - x_*)\|_{Q_\alpha^{-1}}^2 \\
& \quad + (1 - \alpha) \langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle \\
& = \frac{1}{\rho} \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|_{Q_\alpha^{-1}}^2 \\
& \quad - \frac{(1-2\alpha)^2}{\rho-1} \langle x_{n+1} - x_* \mid (\text{Id} - Q)Q_\alpha^{-1}(\text{Id} - Q)(x_{n+1} - x_*) \rangle \\
& \quad + (1 - \alpha) \langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle \\
& \geq \frac{1}{\rho} \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|_{Q_\alpha^{-1}}^2 \tag{59e} \\
& \quad - \frac{(2\alpha-1)^2}{(\rho-1)(1-\alpha)} \langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle \\
& \quad + (1 - \alpha) \langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle \tag{59f} \\
& = \frac{1}{\rho} \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|_{Q_\alpha^{-1}}^2 + \theta \langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle \tag{59g} \\
& \geq \frac{(1-\alpha)}{\rho} \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2 + \theta \langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle, \tag{59h} \\
& \geq \frac{(1-\alpha)}{\rho} \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2 + \theta(1 - \alpha) \|(\text{Id} - Q)(x_{n+1} - x_*)\|_{Q_\alpha^{-1}}^2 \tag{59i} \\
& = \frac{(1-\alpha)}{\rho} \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2 + \theta(1 - \alpha) \|(\text{Id} - T)x_{n+1}\|_{Q_\alpha^{-1}}^2 \tag{59j} \\
& \geq \frac{(1-\alpha)}{\rho} \|Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]\|^2 + \theta(1 - \alpha)^2 \|(\text{Id} - T)x_{n+1}\|^2, \tag{59k}
\end{aligned}$$

where the definition of  $\Upsilon$ , given after (52), was used in (59a), (21) in (59b), (23) in (59c), (26) with  $\eta := \rho/(\rho - 1)$ ,  $a := Uv_{n+1} + \lambda[\nabla f(x_n) + \xi_{n+1}]$ ,  $b := (1 - 2\alpha)(\text{Id} - T)x_{n+1}$  and  $\Pi := Q_\alpha^{-1}$ , as well as  $\rho > 1$  in (59d), and

$$\begin{aligned}
& \langle x_{n+1} - x_* \mid (\text{Id} - Q)Q_\alpha^{-1}(\text{Id} - Q)(x_{n+1} - x_*) \rangle \\
& = \langle x_{n+1} - x_* \mid U^2 Q_\alpha^{-1} U^2 (x_{n+1} - x_*) \rangle \\
& = \langle U(x_{n+1} - x_*) \mid (U Q_\alpha^{-1} U) U(x_{n+1} - x_*) \rangle \\
& \leq \|U Q_\alpha^{-1} U\| \langle U(x_{n+1} - x_*) \mid U(x_{n+1} - x_*) \rangle \tag{60a} \\
& = \|U Q_\alpha^{-1} U\| \langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle \\
& \leq \|U\|^2 \|Q_\alpha^{-1}\| \langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle \\
& = \|\text{Id} - Q\| \|Q_\alpha^{-1}\| \langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle \\
& \leq \frac{1}{1-\alpha} \langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle \tag{60b}
\end{aligned}$$

with (58) and  $\|\text{Id} - Q\| \leq 1$  in (59f). Note that [17, Thm. 9.2-2, p. 466] was used in (60a). Moreover,  $\theta := (1 - \alpha) - (2\alpha - 1)^2/[(1 - \alpha)(\rho - 1)]$  becomes positive for any  $\rho > 1 + (2\alpha - 1)^2/(1 - \alpha)^2$  in (59g), (58) in (59h), (60b) in (59i), the fact  $(\text{Id} - Q)(x_{n+1} - x_*) = (\text{Id} - T)x_{n+1} - (\text{Id} - T)x_* = (\text{Id} - T)x_{n+1}$  in (59j), and (58) in (59k).

Due to (30), the previous considerations suggest that there exists  $C \in \mathbb{R}_{>0}$  s.t.  $\forall n$ ,

$$\begin{aligned}
\frac{C}{n+1} & \geq \frac{1}{n+1} \sum_{\nu=0}^n \|y_{\nu+1} - y_\nu\|_\Theta^2 \\
& \geq \frac{1}{\rho(n+1)} \sum_{\nu=0}^n \|Uv_{\nu+1} + \lambda[\nabla f(x_\nu) + \xi_{\nu+1}]\|^2
\end{aligned}$$

$$\begin{aligned}
& + \frac{\theta}{n+1} \sum_{\nu=0}^n \langle x_{\nu+1} - x_* \mid (\text{Id} - Q)(x_{\nu+1} - x_*) \rangle \\
& \geq \frac{1}{\rho(n+1)} \sum_{\nu=0}^n \|Uv_{\nu+1} + \lambda[\nabla f(x_\nu) + \xi_{\nu+1}]\|^2 + \frac{\theta(1-\alpha)^2}{n+1} \sum_{\nu=0}^n \|(\text{Id} - T)x_{\nu+1}\|^2,
\end{aligned}$$

which establishes the claim of Theorem 20 regarding the sequence of (13). The proof of the claim with regards to the sequence of (48) follows the same steps as the previous one, but with the twist of replacing  $\nabla f(x_n)$  by  $\nabla f(T_\alpha x_n)$  and  $g = 0$ .  $\square$

**Theorem 21.** For the sequence  $(x_n)_{n \in \mathbb{N}}$  of (15), there exists  $\xi_n \in \partial g(x_n)$ ,  $\forall n$ , s.t. for any  $x_* \in \text{Fix } T$ ,

$$\begin{aligned}
\langle x_{n+1} - x_* \mid (\text{Id} - Q)(x_{n+1} - x_*) \rangle &= O\left(\frac{1}{n+1}\right), \\
\|Uv_{n+1} + \lambda\xi_{n+1}\|^2 &= O\left(\frac{1}{n+1}\right), \\
\|(\text{Id} - T)x_{n+1}\|^2 &= O\left(\frac{1}{n+1}\right). \quad \square
\end{aligned}$$

*Proof.* Define here  $\Delta x_n := x_{n-1} - x_n$ ,  $\Delta v_n := v_{n-1} - v_n$ ,  $\Delta y_n := (\Delta x_n, \Delta v_n)$ , and  $\Delta \xi_n := \xi_{n-1} - \xi_n$ ,  $\forall n$ . Under these definitions and in the case of  $f = 0$ , (33) yields

$$\begin{aligned}
& (1 - 2\alpha)(Q - \text{Id})(x_n - x_{n+1}) + Q_\alpha[(x_n - x_{n+1}) - (x_{n-1} - x_n)] \\
& \quad = -U(v_n - v_{n+1}) - \lambda(\xi_n - \xi_{n+1}) \\
& \Leftrightarrow (1 - 2\alpha)(Q - \text{Id})\Delta x_{n+1} + Q_\alpha(\Delta x_{n+1} - \Delta x_n) = -U\Delta v_{n+1} - \lambda\Delta \xi_{n+1} \\
& \Leftrightarrow \lambda\Delta \xi_{n+1} = -U\Delta v_{n+1} - Q_\alpha(\Delta x_{n+1} - \Delta x_n) - (1 - 2\alpha)(Q - \text{Id})\Delta x_{n+1}. \quad (61)
\end{aligned}$$

Moreover, (21) suggests that  $-\Delta v_{n+1} = (1 - \alpha)U(x_{n+1} - x_*)$ , and thus

$$\frac{1}{1-\alpha}(\Delta v_{n+1} - \Delta v_n) = U\Delta x_{n+1}. \quad (62)$$

The monotonicity of  $\partial g(\cdot)$ , (61), (62), and the definition of  $\Theta$ , introduced after (27), imply that

$$\begin{aligned}
& 0 \leq \langle \Delta x_{n+1} \mid \lambda\Delta \xi_{n+1} \rangle \\
& \Leftrightarrow 0 \leq \langle \Delta x_{n+1} \mid -U\Delta v_{n+1} - Q_\alpha(\Delta x_{n+1} - \Delta x_n) - (2\alpha - 1)(\text{Id} - Q)\Delta x_{n+1} \rangle \\
& \Leftrightarrow (2\alpha - 1)\langle \Delta x_{n+1} \mid (\text{Id} - Q)\Delta x_{n+1} \rangle \\
& \quad \leq -\langle U\Delta x_{n+1} \mid \Delta v_{n+1} \rangle - \langle \Delta x_{n+1} \mid Q_\alpha(\Delta x_{n+1} - \Delta x_n) \rangle \\
& \Leftrightarrow (2\alpha - 1)\langle \Delta x_{n+1} \mid (\text{Id} - Q)\Delta x_{n+1} \rangle \\
& \quad \leq -\frac{1}{1-\alpha}\langle \Delta v_{n+1} - \Delta v_n \mid \Delta v_{n+1} \rangle - \langle \Delta x_{n+1} \mid Q_\alpha(\Delta x_{n+1} - \Delta x_n) \rangle \\
& \Leftrightarrow (2\alpha - 1)\langle \Delta x_{n+1} \mid (\text{Id} - Q)\Delta x_{n+1} \rangle \leq \langle \Delta y_{n+1} \mid \Delta y_n - \Delta y_{n+1} \rangle_\Theta \\
& \Leftrightarrow (2\alpha - 1)\langle \Delta x_{n+1} \mid (\text{Id} - Q)\Delta x_{n+1} \rangle \leq \frac{1}{2}(\|\Delta y_n\|_\Theta^2 - \|\Delta y_{n+1}\|_\Theta^2 - \|\Delta y_n - \Delta y_{n+1}\|_\Theta^2) \\
& \Leftrightarrow 2(2\alpha - 1)\langle \Delta x_{n+1} \mid (\text{Id} - Q)\Delta x_{n+1} \rangle + \|\Delta y_n - \Delta y_{n+1}\|_\Theta^2 \\
& \quad \leq \|\Delta y_n\|_\Theta^2 - \|\Delta y_{n+1}\|_\Theta^2, \quad (63)
\end{aligned}$$

and due to  $\alpha \geq 1/2$  as well as the positive-definiteness of  $\text{Id} - Q$ , (63) yields

$$\|y_{n+1} - y_n\|_\Theta^2 \leq \|y_n - y_{n-1}\|_\Theta^2, \quad \forall n. \quad (64)$$

Now, (30) and (64) imply that there exists  $C > 0$  s.t. for any  $n$ ,

$$(n+1)\|y_{n+1} - y_n\|_{\Theta}^2 \leq \sum_{\nu=0}^n \|y_{\nu+1} - y_{\nu}\|_{\Theta}^2 \leq C,$$

and thus  $\|y_{n+1} - y_n\|_{\Theta}^2 \leq C/(n+1)$ . This result applied to (59h) and (59k) establishes the claim of Theorem 21.  $\square$

#### 4. Conclusions

This paper introduced the *Fejér-monotone hybrid steepest descent method* (FM-HSDM) for solving affinely constrained composite minimization tasks in real Hilbert spaces. Only differential and proximal mappings are used to provide low-computational-complexity recursions with enhanced flexibility towards the accommodation of affine constraints. The advocated scheme enjoys Fejér monotonicity, a constant step-size parameter across iterations, and minimal presuppositions on the smooth and non-smooth loss functions to establish weak, and under certain hypotheses, strong convergence to an optimal point. Results on the rate of convergence of the FM-HSDM's sequence of estimates were also presented. Numerical tests on synthetic data were also demonstrated to validate the theoretical findings. Thorough tests on real data, which showcase the flexibility of the family of mappings  $\mathfrak{T}_{\mathcal{A}}$  [cf. (7)] in the context of FM-HSDM, are deferred to an upcoming publication.

#### 5. Acknowledgments

This work was partially supported by the NSF awards 1514056 and 1525194.

#### Appendix A.

Several special cases of  $\mathcal{A}$ , of large interest in optimization tasks, together with members of the family of mappings  $\mathfrak{T}_{\mathcal{A}}$  follow.

**Example 22.** Given a Hilbert space  $\mathfrak{X}_0$  and  $I \in \mathbb{Z}_{>0}$ , consider the Hilbert space  $\mathcal{X} := \mathfrak{X}_0 \times \mathfrak{X}_0 \times \dots \times \mathfrak{X}_0 = \{x := (x^{(1)}, x^{(2)}, \dots, x^{(I)}) \mid x^{(i)} \in \mathfrak{X}_0, \forall i \in \{1, \dots, I\}\}$ , equipped with the inner product  $\langle x \mid x' \rangle_{\mathcal{X}} := \sum_{i=1}^I \langle x^{(i)} \mid x'^{(i)} \rangle$ . Then, upon defining the (closed) linear subspace  $\mathcal{S} := \{x \in \mathcal{X} \mid x^{(1)} = x^{(2)} = \dots = x^{(I)}\}$ , the metric projection mapping onto  $\mathcal{S}$  satisfies

$$P_{\mathcal{S}}(x) = \left( \frac{1}{I} \sum_{i=1}^I x^{(i)}, \frac{1}{I} \sum_{i=1}^I x^{(i)}, \dots, \frac{1}{I} \sum_{i=1}^I x^{(i)} \right), \quad \forall x \in \mathcal{X}, \quad (\text{A.1})$$

and  $P_{\mathcal{S}} \in \mathfrak{T}_{\mathcal{S}}$ .  $\square$

*Proof.* Formula (A.1) can be easily derived by applying Example 7(i) to the special cases of  $\mathcal{X}$  and  $\mathcal{S}$ :  $\|x - P_{\mathcal{S}}x\|_{\mathcal{X}}^2 = \min_{z \in \mathcal{S}} \sum_{i=1}^I \|x^{(i)} - z\|^2$ . Then, claim  $P_{\mathcal{S}} \in \mathfrak{T}_{\mathcal{S}}$  is established by noticing that  $\mathcal{S}$  is a closed affine set and by Proposition 11.  $\square$



**Example 23** (Metric projection mapping onto a hyperplane). For a non-zero  $a \in \mathcal{X}$  and a real number  $b$ , consider the metric projection mapping onto the hyperplane  $\mathcal{H} := \{x \in \mathcal{X} \mid \langle a \mid x \rangle = b\}$  [2, (3.11), p. 49]

$$P_{\mathcal{H}} = \text{Id} - \frac{\langle a \mid \text{Id} \rangle}{\|a\|^2} a + \frac{b}{\|a\|^2} a. \quad (\text{A.2})$$

Then,  $P_{\mathcal{H}} \in \mathfrak{T}_{\mathcal{H}}$ .  $\square$

*Proof.* The claim follows by the observations that  $\mathcal{H}$  is a closed affine set,  $(b/\|a\|^2)a \in \mathcal{H}$ , and by introducing  $\mathcal{V} = \{x \in \mathcal{X} \mid \langle a \mid x \rangle = 0\}$ , with  $P_{\mathcal{V}} = \text{Id} - \frac{\langle a \mid \text{Id} \rangle}{\|a\|^2} a$  and  $P_{\mathcal{V}}[(b/\|a\|^2)a] = 0$ , in Proposition 11.  $\square$

As the following fact states, affine sets obtain a specific form in Euclidean spaces.

**Fact 24** ([20, Thm. 1.4, p. 5]). Given  $\mathbf{b} \in \mathbb{R}^M$  ( $M \in \mathbb{Z}_{>0}$ ) and  $\mathbf{A} \in \mathbb{R}^{M \times D}$  ( $D \in \mathbb{Z}_{>0}$ ) the set  $\{\mathbf{x} \in \mathbb{R}^D \mid \mathbf{A}\mathbf{x} = \mathbf{b}\}$ , if non-empty, is an affine set. Moreover, every affine set in  $\mathcal{X} := \mathbb{R}^D$  can be represented in this way.  $\square$

Motivated by the previous fact and aiming at an algorithmic scheme with wide applicability in Euclidean spaces, where most of the minimization problems reside, the following example and proposition offer a view of affine sets via *least-squares* (LS) tasks and nonexpansive mappings.

**Example 25** (Affinely constrained LS in Euclidean spaces). For vector  $\mathbf{b}$  and matrix  $\mathbf{A}$  of Fact 24, consider the following LS solution set [2, Prop. 3.25, p. 50]:

$$\mathcal{A} := \arg \min_{\mathbf{x} \in \mathbb{R}^D} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = \{\mathbf{x} \in \mathbb{R}^D \mid \mathbf{A}^\top \mathbf{A}\mathbf{x} = \mathbf{A}^\top \mathbf{b}\}. \quad (\text{A.3})$$

Now, considering the  $D \times 1$  vectors  $\{\boldsymbol{\alpha}_m\}_{m=1}^M$ , defined by the rows of  $\mathbf{A}$ , *i.e.*,  $[\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_M] := \mathbf{A}^\top$ , as well as the  $D \times 1$  vectors  $\{\mathbf{g}_d\}_{d=1}^D$  defined via  $[\mathbf{g}_1, \dots, \mathbf{g}_D] := \mathbf{G}$ , where  $\mathbf{G} := \mathbf{A}^\top \mathbf{A}$  and  $\mathbf{c} := [c_1, c_2, \dots, c_D]^\top := \mathbf{A}^\top \mathbf{b}$ , let the hyperplanes  $\mathcal{A}_m := \{\mathbf{x} \in \mathbb{R}^D \mid \langle \boldsymbol{\alpha}_m \mid \mathbf{x} \rangle = b_m\}$ , ( $m = 1, \dots, M$ ), as well as  $\mathcal{G}_d := \{\mathbf{x} \in \mathbb{R}^D \mid \langle \mathbf{g}_d \mid \mathbf{x} \rangle = c_d\}$ , ( $d = 1, \dots, D$ ), with associated metric projection mappings  $P_{\mathcal{A}_m}$  and  $P_{\mathcal{G}_d}$ , respectively [cf. (A.2)]. Then, any of the following mappings, with  $\dagger$  denoting the Moore-Penrose pseudoinverse operation [3],

$$T = \begin{cases} \left( \mathbf{I} - \frac{\mu}{\varrho} \mathbf{A}^\top \mathbf{A} \right) \text{Id} + \frac{\mu}{\varrho} \mathbf{A}^\top \mathbf{b}, & \varrho \geq \|\mathbf{A}\|^2, \mu \in (0, 1], & (\text{A.4a}) \\ (\mathbf{I} - \mathbf{A}^\top \mathbf{A}^\dagger) \text{Id} + \mathbf{A}^\dagger \mathbf{b}, & & (\text{A.4b}) \\ (\mathbf{I} - \mathbf{G} \mathbf{G}^\dagger) \text{Id} + \mathbf{G}^\dagger \mathbf{A}^\top \mathbf{b}, & & (\text{A.4c}) \\ (\mathbf{I} + \gamma \mathbf{A}^\top \mathbf{A})^{-1} \text{Id} + \gamma (\mathbf{I} + \gamma \mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}, & \gamma \in \mathbb{R}_{>0}, & (\text{A.4d}) \\ (1 - \beta) \text{Id} + \beta \sum_{m=1}^M \frac{\|\boldsymbol{\alpha}_m\|^2}{\|\mathbf{A}\|_{\text{F}}^2} P_{\mathcal{A}_m}, & \beta \in (0, 1], & (\text{A.4e}) \\ (1 - \theta) \text{Id} + \theta \sum_{d=1}^D \omega_d P_{\mathcal{G}_d}, & \begin{cases} \theta \in (0, 1], \omega_d \in (0, 1), \\ \sum_{d=1}^D \omega_d = 1, \end{cases} & (\text{A.4f}) \end{cases}$$

satisfies  $T \in \mathfrak{T}_{\mathcal{A}}$ .

Further, given also the  $M_0 \times 1$  ( $M_0 \in \mathbb{Z}_{>0}$ ) vector  $\mathbf{b}_0$ , the  $M_0 \times D$  matrix  $\mathbf{A}_0$ , let the non-empty affine constraint set  $\mathcal{K} := \{\mathbf{x} \in \mathbb{R}^D \mid \mathbf{A}_0 \mathbf{x} = \mathbf{b}_0\}$ , with metric projection mapping  $P_{\mathcal{K}} = (\mathbf{I} - \mathbf{A}_0^\top \mathbf{A}_0^{\dagger\top}) \text{Id} + \mathbf{A}_0^\dagger \mathbf{b}_0$  [2, Prop. 3.17, p. 47]. Then, according to [3, Ex. 34, p. 120],

$$\begin{aligned} \mathbf{x} \in \mathcal{A}_{\mathcal{K}} &:= \arg \min_{\mathbf{z} \in \mathcal{K}} \frac{1}{2} \|\mathbf{A}\mathbf{z} - \mathbf{b}\|^2 \\ \Leftrightarrow \exists \boldsymbol{\mu} \in \mathbb{R}^{M_0} \text{ s.t. } (\mathbf{x}, \boldsymbol{\mu}) \in \overline{\mathcal{A}} &:= \left\{ (\mathbf{x}', \boldsymbol{\mu}') \in \mathbb{R}^D \times \mathbb{R}^{M_0} \left| \overbrace{\begin{bmatrix} \mathbf{A}^\top \mathbf{A} & \mathbf{A}_0^\top \\ \mathbf{A}_0 & \mathbf{0} \end{bmatrix}}^{\mathbf{L} :=} \begin{bmatrix} \mathbf{x}' \\ \boldsymbol{\mu}' \end{bmatrix} = \overbrace{\begin{bmatrix} \mathbf{A}^\top \mathbf{b} \\ \mathbf{b}_0 \end{bmatrix}}^{\mathbf{e} :=} \right\}, \end{aligned} \quad (\text{A.5})$$

or, in other words,  $\mathcal{A}_{\mathcal{K}} = \Pi_{\mathbb{R}^D} \overline{\mathcal{A}}$ , where  $\Pi_{\mathbb{R}^D}$  denotes the mapping  $\Pi_{\mathbb{R}^D} : \mathbb{R}^D \times \mathbb{R}^{M_0} \rightarrow \mathbb{R}^D : (\mathbf{x}, \boldsymbol{\mu}) \mapsto \mathbf{x}$ . Define also the  $(D + M_0) \times 1$  vectors  $[\mathbf{l}_1, \dots, \mathbf{l}_{D+M_0}] := \mathbf{L}$ , as well as the hyperplanes  $\mathcal{L}_d := \{(\mathbf{x}', \boldsymbol{\mu}') \in \mathbb{R}^D \times \mathbb{R}^{M_0} \mid \langle \mathbf{l}_d \mid (\mathbf{x}', \boldsymbol{\mu}') \rangle = e_d\}$ , with  $P_{\mathcal{L}_d}$  denoting the associated metric projection mapping [cf. (A.2)]. Then, any of the following mappings  $\overline{T} : \mathbb{R}^{D+M_0} \rightarrow \mathbb{R}^{D+M_0}$ :

$$\overline{T} = \begin{cases} \left( \mathbf{I} - \frac{\bar{\varrho}}{\varrho} \mathbf{L}^\top \mathbf{L} \right) \text{Id} + \frac{\bar{\varrho}}{\varrho} \mathbf{L}^\top \mathbf{b}, & \bar{\varrho} \geq \|\mathbf{L}\|^2, \bar{\mu} \in (0, 1], & (\text{A.6a}) \\ \left( \mathbf{I} - \mathbf{L}^\top \mathbf{L}^{\dagger\top} \right) \text{Id} + \mathbf{L}^\dagger \mathbf{e}, & & (\text{A.6b}) \\ \left( \mathbf{I} + \bar{\gamma} \mathbf{L}^\top \mathbf{L} \right)^{-1} \text{Id} + \bar{\gamma} \left( \mathbf{I} + \bar{\gamma} \mathbf{L}^\top \mathbf{L} \right)^{-1} \mathbf{L}^\top \mathbf{e}, & \bar{\gamma} \in \mathbb{R}_{>0}, & (\text{A.6c}) \\ \left( 1 - \bar{\theta} \right) \text{Id} + \bar{\theta} \sum_{d=1}^{D+M_0} \bar{w}_d P_{\mathcal{L}_d}, & \begin{cases} \bar{\theta} \in (0, 1], \bar{w}_d \in (0, 1), \\ \sum_{d=1}^{D+M_0} \bar{w}_d = 1, \end{cases} & (\text{A.6d}) \end{cases}$$

satisfies  $\overline{T} \in \mathfrak{T}_{\overline{\mathcal{A}}}$ . Moreover, the mapping  $T : \mathbb{R}^D \rightarrow \mathbb{R}^D$ , defined by

$$T := (1 - \bar{\beta}) P_{\mathcal{K}} + \bar{\beta} P_{\mathcal{K}} \sum_{m=1}^M \frac{\|\boldsymbol{\alpha}_m\|^2}{\|\mathbf{A}\|_{\text{F}}^2} P_{\mathcal{A}_m} P_{\mathcal{K}}, \quad \bar{\beta} \in (0, 1], \quad (\text{A.6e})$$

satisfies  $T \in \mathfrak{T}_{\mathcal{A}_{\mathcal{K}}}$ . □

*Proof.* For  $\delta \in \mathbb{R}_{>0}$ , define

$$\varphi_\delta(\mathbf{x}) := \frac{1}{2\delta} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2, \quad \forall \mathbf{x} \in \mathbb{R}^D, \quad (\text{A.7})$$

and verify that  $\nabla \varphi_\delta = (1/\delta) \mathbf{A}^\top \mathbf{A} \text{Id} - (1/\delta) \mathbf{A}^\top \mathbf{b}$ . According to (A.3), all points  $\mathbf{x} \in \mathbb{R}^D$  s.t.  $\nabla \varphi_\delta(\mathbf{x}) = \mathbf{0}$  constitute  $\mathcal{A}$ . Moreover, for any  $\varrho \geq \|\mathbf{A}\|^2/\delta$ ,  $\|\nabla \varphi_\delta(\mathbf{x}) - \nabla \varphi_\delta(\mathbf{x}')\| \leq (\|\mathbf{A}\|^2/\delta) \|\mathbf{x} - \mathbf{x}'\| \leq \varrho \|\mathbf{x} - \mathbf{x}'\|$ ,  $\forall \mathbf{x}, \mathbf{x}' \in \mathbb{R}^D$ , since  $\|\mathbf{A}^\top \mathbf{A}\| = \|\mathbf{A}\|^2$ . In other words,  $\nabla \varphi_\delta$  is  $\varrho$ -Lipschitz continuous, which, according to the Baillon-Haddad theorem [1], [2, Cor. 18.16, p. 270], is equivalent to that  $(1/\varrho) \nabla \varphi_\delta$  is firmly nonexpansive iff  $\text{Id} - (1/\varrho) \nabla \varphi_\delta$  is firmly nonexpansive [cf. Example 7(iii)] with fixed-point set equal to  $\mathcal{A}$ . By utilizing once again Example 7(iii),  $R := 2[\text{Id} - (1/\varrho) \nabla \varphi_\delta] - \text{Id}$  is nonexpansive, and for any  $\zeta \in (0, 1]$ ,  $R' := \zeta R + (1 - \zeta) \text{Id} = \text{Id} - (2\zeta/\varrho) \nabla \varphi_\delta = \{\mathbf{I} - [2\zeta/(\varrho\delta)] \mathbf{A}^\top \mathbf{A}\} \text{Id} + [2\zeta/(\varrho\delta)] \mathbf{A}^\top \mathbf{b}$  is nonexpansive with  $\text{Fix}(R') = \mathcal{A}$ . Due to the nonexpansiveness of  $R'$ ,  $\|\mathbf{I} - [2\zeta/(\varrho\delta)] \mathbf{A}^\top \mathbf{A}\| \leq 1$  (cf. Fact 9). Constraining  $\zeta \in (0, 1/2]$

guarantees that  $\mathbf{I} - [2\zeta/(\varrho\delta)]\mathbf{A}^\top\mathbf{A} \succeq \mathbf{0}$ . By defining  $\mu := 2\zeta$  and  $\delta := 1$ , the claim regarding (A.4a) is established.

The metric projection mapping  $P_{\ker \mathbf{A}}$  onto  $\ker \mathbf{A}$  is  $P_{\ker \mathbf{A}} = (\mathbf{I} - \mathbf{A}^\top \mathbf{A}^{\top\dagger}) \text{Id}$  [2, Prop. 3.28(iii), p. 51]. Since  $\mathcal{A} = \ker \mathbf{A} + \mathbf{A}^\dagger \mathbf{b}$  [2, Prop. 3.28(i), p. 51], [2, Prop. 3.17, p. 47] suggests that the metric projection mapping  $P_{\mathcal{A}}$  onto  $\mathcal{A}$  becomes  $P_{\mathcal{A}} = P_{\ker \mathbf{A}} + \mathbf{A}^\dagger \mathbf{b} - P_{\ker \mathbf{A}}(\mathbf{A}^\dagger \mathbf{b}) = P_{\ker \mathbf{A}} + \mathbf{A}^\dagger \mathbf{b}$ , due to  $P_{\ker \mathbf{A}}(\mathbf{A}^\dagger \mathbf{b}) = \mathbf{0}$  [2, Prop. 3.28(i), p. 51]. Hence, (A.4b) is an immediate consequence of Proposition 11. By [3, Ex. 18(d), p. 49],  $\mathbf{A}^\top \mathbf{A}^{\top\dagger} = \mathbf{A}^\top \mathbf{A}(\mathbf{A}^\top \mathbf{A})^\dagger = \mathbf{G}\mathbf{G}^\dagger$  and  $\mathbf{A}^\dagger \mathbf{b} = (\mathbf{A}^\top \mathbf{A})^\dagger \mathbf{A}^\top \mathbf{b} = \mathbf{G}^\dagger \mathbf{A}^\top \mathbf{b}$ . Hence, (A.4c) follows easily from (A.4b).

Now, for any  $\gamma' \in \mathbb{R}_{>0}$ ,  $\text{Prox}_{\gamma'\varphi_\delta} = (\mathbf{I} + (\gamma'/\delta)\mathbf{A}^\top \mathbf{A})^{-1} \text{Id} + (\gamma'/\delta)(\mathbf{I} + (\gamma'/\delta)\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b}$ . Setting  $\gamma := \gamma'/\delta$ , the nonexpansiveness of  $\text{Prox}_{\gamma\delta\varphi_\delta}$ , stated by Example 7(ii), suggests that  $\|(\mathbf{I} + \gamma\mathbf{A}^\top \mathbf{A})^{-1}\| \leq 1$  (cf. Fact 9), and that  $\text{Fix}(\text{Prox}_{\gamma\delta\varphi_\delta}) = \mathcal{A}$ . Due also to the fact that  $(\mathbf{I} + \gamma\mathbf{A}^\top \mathbf{A})^{-1}$  is positive, the claim regarding (A.4d) is established.

Let  $\delta := \|\mathbf{A}\|_{\text{F}}^2$  in (A.7), so that

$$\begin{aligned} \varphi_{\|\mathbf{A}\|_{\text{F}}^2}(\mathbf{x}) &= \frac{1}{2\|\mathbf{A}\|_{\text{F}}^2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = \frac{1}{2\|\mathbf{A}\|_{\text{F}}^2} \sum_{m=1}^M (\langle \boldsymbol{\alpha}_m | \mathbf{x} \rangle - b_m)^2 \\ &= \frac{1}{2} \sum_{m=1}^M \frac{\|\boldsymbol{\alpha}_m\|^2}{\|\mathbf{A}\|_{\text{F}}^2} \|\mathbf{x} - P_{\mathcal{A}_m}(\mathbf{x})\|^2 = \frac{1}{2} \sum_{m=1}^M w_m \|\mathbf{x} - P_{\mathcal{A}_m}(\mathbf{x})\|^2, \end{aligned}$$

where the explicit expression of  $P_{\mathcal{A}_m}$  is given in (A.2), and the non-negative weights  $\{w_m := \|\boldsymbol{\alpha}_m\|^2 / \|\mathbf{A}\|_{\text{F}}^2\}_{m=1}^M$  satisfy  $\sum_{m=1}^M w_m = 1$ . It can be also verified by the Fréchet-gradient definition [2, Def. 2.45, p. 38] that  $\nabla\|(\text{Id} - P_{\mathcal{A}_m})\mathbf{x}\|^2 = 2(\text{Id} - P_{\mathcal{A}_m})\mathbf{x}$ , which yields

$$\nabla\varphi_{\|\mathbf{A}\|_{\text{F}}^2} = \sum_{m=1}^M w_m (\text{Id} - P_{\mathcal{A}_m}) = \text{Id} - \sum_{m=1}^M w_m P_{\mathcal{A}_m}.$$

Hence, all minimizers of  $\varphi_{\|\mathbf{A}\|_{\text{F}}^2}$ , i.e.,  $\mathcal{A}$ , constitute the fixed-point set of  $\sum_m w_m P_{\mathcal{A}_m}$ , which is equal to the fixed-point set of the mapping in (A.4e). Hence, by utilizing the trivial fact  $\text{Id} \in \mathfrak{T}$  and by applying also Proposition 10(i) to  $(1 - \beta)\text{Id} + \beta \sum_m w_m P_{\mathcal{A}_m}$ , the claim of (A.4e) is established.

Regarding (A.4f), notice first that  $\mathcal{A} = \cap_{d=1}^D \mathcal{G}_d$ . According to Example 7(iv),  $\mathcal{A} = \text{Fix}(\sum_d \omega_d P_{\mathcal{G}_d})$ . Since  $P_{\mathcal{G}_d} \in \mathfrak{T}$  (cf. Example 23), Proposition 10(i) yields  $\sum_d \omega_d P_{\mathcal{G}_d} \in \mathfrak{T}$ . As a result, fact  $\text{Id} \in \mathfrak{T}$  and Proposition 10(i) yield  $(1 - \theta)\text{Id} + \theta \sum_d \omega_d P_{\mathcal{G}_d} \in \mathfrak{T}$ , which establishes the claim of (A.4f). Due to  $\overline{\mathcal{A}} = \arg \min_{(\mathbf{x}, \boldsymbol{\mu})} \|\mathbf{L}[\mathbf{x}^\top, \boldsymbol{\mu}^\top]^\top - \mathbf{e}\|^2$ , arguments similar to those developed for (A.4a), (A.4b) and (A.4d) yield (A.6a), (A.6b) and (A.6c), respectively. Further, notice that since  $\overline{\mathcal{A}} = \cap_{d=1}^{D+M_0} \mathcal{L}_d$ , (A.6d) is deduced in a way similar to the derivation of (A.4f) from (A.3).

Regarding (A.6e), notice that  $\mathcal{A}_{\mathcal{K}} = \text{Fix } T_{\mathcal{A}_{\mathcal{K}}}$  [26, Prop. 4.2(a)], where

$$T_{\mathcal{A}_{\mathcal{K}}} := (1 - \bar{\beta})\text{Id} + \bar{\beta} P_{\mathcal{K}} \sum_{m=1}^M \frac{\|\boldsymbol{\alpha}_m\|^2}{\|\mathbf{A}\|_{\text{F}}^2} P_{\mathcal{A}_m}$$

is nonexpansive for  $\bar{\beta} \in (0, 3/2]$ . Since  $\mathcal{A}_{\mathcal{K}} = \text{Fix } T_{\mathcal{A}_{\mathcal{K}}} = \text{Fix } T_{\mathcal{A}_{\mathcal{K}}} \cap \mathcal{K} = \text{Fix } T_{\mathcal{A}_{\mathcal{K}}} \cap \text{Fix } P_{\mathcal{K}}$ , Example 7(v) suggests that  $\mathcal{A}_{\mathcal{K}}$  can be seen also as the fixed-point set of

the nonexpansive mapping  $T_{\mathcal{A}_K}P_K$ , which is nothing but the mapping appearing at (A.6e). Now, due to Proposition 10(i) and Example 23,  $\sum_m w_m P_{\mathcal{A}_m} \in \mathfrak{T}$ , with  $w_m := \|\alpha_m\|^2 / \|\mathbf{A}\|_{\mathbb{F}}^2$ . Hence, Proposition 10(ii) suggests also that  $P_K(\sum_m w_m P_{\mathcal{A}_m})P_K \in \mathfrak{T}$ . Once again, since  $P_K \in \mathfrak{T}$  (cf. Proposition 11), Proposition 10(i) guarantees  $(1 - \bar{\beta})P_K + \bar{\beta}P_K \sum_m w_m P_{\mathcal{A}_m}P_K \in \mathfrak{T}$ , for  $\bar{\beta} \in (0, 1]$ , which establishes the claim of (A.6e).  $\square$

An auxiliary proposition, used in Theorem 20, follows.

**Proposition 26.** Given the surjective and strongly positive mapping  $\Pi \in \mathfrak{B}(\mathcal{X})$ , i.e., there exists  $\delta \in \mathbb{R}_{>0}$  s.t.  $\langle \Pi x \mid x \rangle \geq \delta \|x\|^2$ ,  $\forall x \in \mathcal{X}$ , the inverse  $\Pi^{-1}$  exists and  $\Pi^{-1} \in \mathfrak{B}(\mathcal{X})$  with  $\|\Pi^{-1}\| \leq 1/\delta$ . Moreover,  $\Pi^{-1}$  is strongly positive and  $(\delta/\|\Pi\|^2)\|x\|^2 \leq \langle \Pi^{-1}x \mid x \rangle \leq (1/\delta)\|x\|^2$ ,  $\forall x \in \mathcal{X}$ .  $\square$

*Proof.* [17, §2.7, Prob. 7, p. 101] guarantees the existence of  $\Pi^{-1}$  and  $\Pi^{-1} \in \mathfrak{B}(\mathcal{X})$ . By the strong positiveness of  $\Pi$ ,  $\forall x \in \mathcal{X} \setminus (\{0\} = \ker \Pi^{-1})$ ,  $\|\Pi^{-1}x\|^2 \leq (1/\delta)\langle \Pi^{-1}x \mid \Pi(\Pi^{-1}x) \rangle = (1/\delta)\langle \Pi^{-1}x \mid x \rangle \leq (1/\delta)\|\Pi^{-1}x\|\|x\| \Rightarrow \|\Pi^{-1}x\| \leq (1/\delta)\|x\| \Rightarrow \|\Pi^{-1}\| \leq (1/\delta)$ . By [17, Thm. 9.4-2, p. 476] and the previous result,  $\forall x \in \mathcal{X}$ ,  $\langle \Pi^{-1}x \mid x \rangle \leq \|\Pi^{-1}\|\|x\|^2 \leq (1/\delta)\|x\|^2$ . Moreover,  $\forall x' \in \mathcal{X}$ ,  $\langle \Pi x' \mid \Pi^{-1}\Pi x' \rangle = \langle \Pi x' \mid x' \rangle \geq \delta \|x'\|^2 \geq (\delta/\|\Pi\|^2)\|\Pi x'\|^2$ , which yields, under  $x := \Pi x'$ , that  $\forall x \in \mathcal{X}$ ,  $(\delta/\|\Pi\|^2)\|x\|^2 \leq \langle \Pi^{-1}x \mid x \rangle$ .  $\square$

## References

- [1] J.-B. Baillon and G. Haddad, “Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones,” *Israel J. of Math.*, vol. 26, no. 2, pp. 137–150, 1977.
- [2] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York: Springer, 2011.
- [3] A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, 2nd ed. New York: Springer-Verlag, 2003.
- [4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [5] K. Bredies and H. Sun, “Preconditioned Douglas-Rachford splitting methods for convex-concave saddle-point problems,” *SIAM J. Numerical Analysis*, vol. 53, no. 1, pp. 421–444, 2015.
- [6] V. Cevher, C. B. Vü, and A. Yurtsever, “Stochastic forward-Douglas-Rachford splitting for monotone inclusions,” EPFL, Tech. Rep. 215759, 2016.
- [7] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *J. Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.
- [8] P. L. Combettes and I. Yamada, “Compositions and convex combinations of averaged nonexpansive operators,” *J. Math. Anal. Appl.*, vol. 425, pp. 55–70, 2015.
- [9] P. L. Combettes and J.-C. Pesquet, “Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators,” *Set-Valued and Variational Analysis*, vol. 20, no. 2, pp. 307–330, 2012.
- [10] L. Condat, “A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,” *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.

- [11] D. Davis and W. Yin, “A three-operator splitting scheme and its optimization applications,” *eprint arXiv:1504.01032v1*, 2015.
- [12] D. Gabay and B. Mercier, “A dual algorithm for the solution of nonlinear variational problems via finite-element approximations,” *Comp. Math. Appl.*, vol. 2, pp. 17–40, 1976.
- [13] R. Glowinski and A. Marrocco, “Sur l’approximation par éléments finis et la résolution par pénalisation-dualité d’une classe de problèmes de Dirichlet non linéaires,” *Rev. Francaise d’Aut. Inf. Rech. Oper.*, vol. 9, no. 2, pp. 41–76, 1975.
- [14] H. Iiduka, “Three-term conjugate gradient method for the convex optimization problem over the fixed point set of a nonexpansive mapping,” *Applied Math. Computation*, vol. 217, pp. 6315–6327, 2011.
- [15] —, “Acceleration method for convex optimization over the fixed point set of a nonexpansive mapping,” *Math. Program.*, vol. 149, pp. 131–165, 2015.
- [16] H. Iiduka and I. Yamada, “A use of conjugate gradient direction for the convex optimization problem over the fixed point set of a nonexpansive mapping,” *SIAM J. Optim.*, vol. 19, no. 4, pp. 1881–1893, 2009.
- [17] E. Kreyszig, *Introductory Functional Analysis and Applications*, Wiley Classics Library ed. New York: Wiley, 1989.
- [18] I. Necoara and A. Patrascu, “A random coordinate descent algorithm for optimization problems with composite objective function and linear coupled constraints,” *Computational Optim. Appl.*, vol. 57, no. 2, pp. 307–337, 2014.
- [19] N. Ogura and I. Yamada, “Non-strictly convex minimization over the fixed point set of the asymptotically shrinking nonexpansive mapping,” *Numerical Functional Analysis and Optimization*, vol. 23, pp. 113–137, 2002.
- [20] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton University Press, 1970.
- [21] W. Shi, Q. Ling, G. Wu, and W. Yin, “EXTRA: An exact first-order algorithm for decentralized consensus optimization,” *SIAM J. Optim.*, vol. 25, no. 2, 2014.
- [22] —, “A proximal gradient algorithm for decentralized composite optimization,” *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6013–6023, 2015.
- [23] K. Slavakis and I. Yamada, “Accelerated hybrid steepest descent method for solving affinely constrained convex composite optimization problems,” in *Intern. Conf. Continuous Optim. (ICCOPT)*, Tokyo: Japan, Aug. 2016.
- [24] K. Slavakis, I. Yamada, and S. Ono, “Accelerating the hybrid steepest descent method for affinely constrained convex composite minimization tasks,” in *Proc. ICASSP*, New Orleans: USA, Mar. 2017.
- [25] H. Sun, “Analysis of fully preconditioned ADMM with relaxation in Hilbert spaces,” *eprint arXiv:1611.04801v2*, 2016.
- [26] I. Yamada, “The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings,” in *Inherently Parallel Algorithms for Feasibility and Optimization and their Applications*, D. Butnariu, Y. Censor, and S. Reich, Eds. Elsevier, 2001, pp. 473–504.